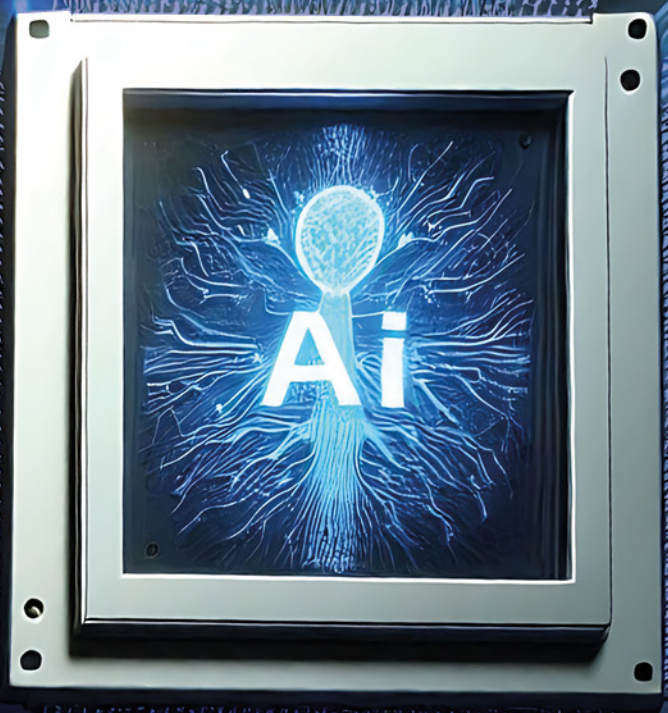
INADI
Instituto para el Desarrollo Industrial y la Transformación Digital A.C.

DPL Group

INTELIGENCIA ARTIFICIAL



Hacia una nueva era en la historia de la humanidad

Arturo Oropeza García
Jorge Fernando Negrete

coordinadores

INADI
Instituto para el Desarrollo Industrial y la Transformación Digital A.C.

DPL Group

INTELIGENCIA ARTIFICIAL

Hacia una nueva era en la historia de la humanidad

Carlos Baigorri * Julen Berasaluce Iza
Fausto Carbajal * Rosario Castañón
Carlos A. Coello Coello * Amaury Fierro González
Mariel García Montes * David Guillén
Zhou Hui * Badar Alam Iqbal
Song Jin * Beatriz Juárez
Carlos López Blanco * José Ramón López-Portillo Romano
Malena Mancini * Sebastian Porsdam Mann
Timo Minssen * Jorge Fernando Negrete
Arturo Oropeza García * Clemente Ruiz Durán
Saiph Savage * Claudia Schatan
Hugo Simg * Hannah Louise Smith
José Luis Solleiro * Osamu Sudoh
Kiyoshi Tsuru * Maria L. Vázquez
Arti Yadav * Amikam Yalovetzky
Pang Zhongying * Mohamad Zreik

Arturo Oropeza García
Jorge Fernando Negrete

Coordinadores

INADI
Instituto para el Desarrollo Industrial y la Transformación Digital A.C.

Inteligencia Artificial. Hacia una nueva era en la historia de la humanidad.
Coordinadores: Arturo Oropeza García y Jorge Fernando Negrete

Primera edición, 2025.

©2025, Instituto para el Desarrollo Industrial y la Transformación Digital A.C. (INADI)

©2025, Arturo Oropeza García
Coyoacán, Ciudad de México

ISBN: 978-607-99893-1-6

Todos los Derechos reservados.

Las opiniones expresadas en este libro son responsabilidad exclusiva de los autores y no reflejan necesariamente la posición de las instituciones a las que pertenecen.

Queda rigurosamente prohibida, sin autorización escrita del titular del derecho, la reproducción o distribución total o parcial de esta obra por cualquier medio o procedimiento bajo sanciones establecidas por la ley.

Colaboración editorial: Alfonso Coronel Urrieta
Diseño editorial: alexbrije

Imágenes de portada e interiores: ©Freepik
Impreso y hecho en México/ Printed and made in Mexico

INADI
Instituto para el Desarrollo Industrial y la Transformación Digital A.C.

DPL Group

INTELIGENCIA ARTIFICIAL

Hacia una nueva era en la historia de la humanidad

Arturo Oropeza García
Jorge Fernando Negrete

coordinadores

ÍNDICE

INTELIGENCIA ARTIFICIAL HACIA UNA NUEVA ERA EN LA HISTORIA DE LA HUMANIDAD

Introducción
> ARTURO OROPEZA GARCÍA
> JORGE FERNANDO NEGRETE

CAPÍTULO I
INTELIGENCIA ARTIFICIAL. HACIA UNA NUEVA ERA DE LA HUMANIDAD

1. Gobernanza de la inteligencia artificial: hacia una nueva era de responsabilidad y bien común
> JOSÉ RAMÓN LÓPEZ-PORTILLO ROMANO
Profesor de Oxford, cofundador del Centro de Estudios Mexicanos y miembro del Consejo Científico Internacional

2. Liberalismo, democracia e inteligencia artificial. Los retos de la libertad
> ARTURO OROPEZA GARCÍA
Presidente del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI) e Investigador Nacional del Instituto de Investigaciones Jurídicas de la UNAM

3. Inteligencia artificial: entre la disrupción y la regulación
> JORGE FERNANDO NEGRETE
Presidente de Digital Policy & Law Group

CAPÍTULO II
INTELIGENCIA ARTIFICIAL. LÍNEAS GENERALES DE UN NUEVO CONOCIMIENTO

4. Pasado, presente y futuro de la inteligencia artificial
> CARLOS A. COELLO COELLO
Miembro del Colegio Nacional. Investigador en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN)

5. La era de la inteligencia artificial y los desafíos en América Latina
> AMIKAM YALOVETZKY (Israel)
Senior Manager & Corporate Marketing Communications en MediaTek
> HUGO SIMG
Latin America Business Development Director en MediaTek

CAPÍTULO III
INTELIGENCIA ARTIFICIAL. LAS NUEVAS POSIBILIDADES DEL DESARROLLO ECONÓMICO

6. Futuros imaginados, expectativas ficcionales y dinámica del capitalismo; el papel de la inteligencia artificial
> CLEMENTE RUIZ DURÁN
Excoordinador del Posgrado en Economía, UNAM. Consejero del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI)

7. Inteligencia artificial y las líneas de construcción de una nueva economía digital
> JULEN BERASALUCE IZA
Profesor-Investigador del Centro de Estudios Económicos de El Colegio de México y Consejero del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI)

CAPÍTULO IV
INTELIGENCIA ARTIFICIAL. SU RELEVANCIA EN EL NUEVO DESARROLLO INDUSTRIAL DIGITAL

8. Estrategia para la adopción de la IA en la industria mexicana 241

> JOSÉ LUIS SOLLEIRO
Investigador Titular del Instituto de Ciencias Aplicadas y Tecnología (ICAT) de la UNAM y Consejero del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI)

> ROSARIO CASTAÑÓN
Académica del Instituto de Ciencias Aplicadas y Tecnología (ICAT) de la UNAM

> DAVID GUILLÉN
Gerente de consultoría y Gestión de Proyectos en Cambiotec, A.C.

9. Impacto de la inteligencia artificial en la industria manufacturera y el caso de México 266

> CLAUDIA SCHATAN
Especialista en temas económicos y Consejera del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI)

10. Inteligencia artificial industrial en China 296

> BEATRIZ JUÁREZ
Profesora de la División de Estudios de Posgrado de la Facultad de Derecho de la UNAM y Consejera del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI)

CAPÍTULO V
INTELIGENCIA ARTIFICIAL. EL SURGIMIENTO DE UNA NUEVA ERA EN MATERIA DE REGULACIÓN

11. El despegue de la inteligencia artificial en el caso brasileño. Avances y retos regulatorios por cumplir 317

> CARLOS BAIGORRI (Brasil)
Presidente y CEO de la Agencia Nacional de Telecomunicaciones de Brasil (ANATEL)

12. The European AI Act: Promises & perils. Trade offs in a complex regulatory eco-system 330

> TIMO MINSSEN (Dinamarca)
Profesor de Derecho y Director del Centro de Estudios Avanzados en Derecho de la Innovación en Biociencia (CeBIL) en la Universidad de Copenhague

> HANNAH LOUISE SMITH (Dinamarca)
Investigadora postdoctoral en el Centro de Estudios Avanzados en Derecho de la Innovación Biomédica

> SEBASTIAN PORSDAM MANN (Dinamarca)
Profesor Asistente en el el Centro de Estudios Avanzados en Derecho de la Innovación Biomédica

13. El futuro del derecho de autor en América Latina en la era de la inteligencia artificial 372

> MARIA L. VÁZQUEZ (Argentina)
Decana de la Escuela de Derecho de la Universidad de San Andrés (UdeSA, Buenos Aires, Argentina). Directora de Maestría de la Propiedad Intelectual e Innovación organizada en conjunto con la Organización Mundial de la Propiedad Intelectual (OMPI)

> MALENA MANCINI (Argentina)
Abogada por la Universidad de Buenos Aires (UBA), especializada en propiedad intelectual y derecho del Entretenimiento

14. General Framework of Regulation in the People's Republic of China on Artificial Intelligence and Digital Transformation 426

> ZHOU HUI (China)
Subdirector, profesor asociado. Departamento de Derecho Cibernético e Informático, Instituto de Derecho de la Academia China de Ciencias Sociales (CASS)

> YUE ZHU (China)
Profesor asistente en Derecho por la Universidad de Tongji

15. Inteligencia artificial. ¿A qué objetivos debe responder la regulación de la inteligencia artificial en México? 448
> MARIEL GARCÍA MONTES
Investigadora y candidata a doctorado en Ciencia, Tecnología y Sociedad, MIT
> SAIPH SAVAGE
Profesora asistente en la Universidad Northeastern, en el Khoury College of Computer Sciences

16. La Inteligencia Artificial Generativa y su impacto en la creatividad y en el Derecho Intelectual 470
> KIYOSHI TSURU
Socio fundador y director de TMI Law. Kiyoshi es un abogado líder especializado en propiedad intelectual. Country Manager de BSA | The Software Alliance. Representante en México de la Entertainment Software Association (ESA).

CAPÍTULO VI
INTELIGENCIA ARTIFICIAL.
UNA VISIÓN GLOBAL COMPARADA 494

17. La inteligencia artificial en el mundo digital del siglo XXI. Una visión europea y latinoamericana 495
> CARLOS LÓPEZ BLANCO (España)
Patrono de la Fundación Euroamérica. Presidente de la Comisión de Economía Digital de la International Chamber of Commerce

18. The Age of AI in China: a partially consideration on some neglected issues 524
> PANG ZHONGYING (China)
Académico en humanidades y ciencias sociales, enseña Relaciones Internacionales Globales, Economía Política Internacional e Historia Económica Global en la Universidad de Sichuan en Chengdu, China

19. The impact of AI on China's economic development. Advances and challenges to be adressed 542
> SONG JIN (China)
Directora del Centro de Investigación en Desarrollo del Instituto de Economía y Política Mundial (IWEP) e investigador de la Academia China de Ciencias Sociales (CASS)

20. Advanced Generative artificial intelligence and AI governance in Japan 562
> OSAMU SUDOH (Japón)
Profesor en la Universidad Chuo

21. Artificial Intelligence in India: Advances and pending challenges 596
> BADAR ALAM IQBAL (India)
Investigador, Departamento de Economía, Universidad de Stellenbosch, Sudáfrica
> MOHAMAD ZREIK (India)
Investigador postdoctoral, Escuela de Estudios Internacionales, Universidad Sun Yat-sen
> ARTI YADAV (India)
Profesor asistente, Departamento de Comercio, Ramanujan College, Universidad de Delhi, India

CAPÍTULO VII
INTELIGENCIA ARTIFICIAL. LOS NUEVOS POTENCIALES EN SUPERVISIÓN Y SEGURIDAD PÚBLICA 618

22. Inteligencia artificial, desinformación y seguridad pública en América Latina 619
> FAUSTO CARBAJAL
Senior Policy Advisor en Miranda Partners, así como socio fundador de Delphi Solutions and Strategic Services y y Consejero del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI)
> AMAURY FIERRO GONZÁLEZ
Especialista en inteligencia y geopolítica e instructor en Rabdan Academy

Autores 648

INTRODUCCIÓN

I.

La historia del mundo desbordada de épicas y retos fue cimbrada a mediados del siglo XVIII cuando un terremoto civilizatorio se presentó en Europa a través de una Primera Revolución Industrial, que vino a modificar la vida del ser humano en su entorno personal, familiar, social, económico y político, para siempre.

A partir de esa primera revolución científica el ser humano no ha vuelto a saber lo que es el reposo.

A principios del siglo XIX, como resultado de ese cambio de vida la población mundial se expandió velozmente alcanzando la cifra de mil millones de seres humanos. La producción económica se disparó; nació lo que sería un nuevo orden económico del mundo; de igual modo que se sucedieron los primeros registros de distribución del ingreso con base a una nueva era de la globalización de la manufactura y del comercio facilitada por la tecnología.

Desde la mitad del siglo XVIII y a lo largo de los siglos XIX y XX, la mayoría de la población aprendió a vivir entre la fascinación de las innovaciones que día a día la deslumbraban, en combinación con el temor de un presente y un futuro que no acababa de entender y que tampoco sabía a dónde la iban a llevar.

En la Exposición Universal de París de 1900, como ejemplo de lo anterior, los visitantes caminaban deslumbrados a través de los largos pasillos de los Jardines del Trocadero hasta la Torre Eiffel, contemplando con gran asombro el telégrafo sin hilos, el telescopio más poderoso del mundo, la primera escalera mecánica, el cinematógrafo, los automóviles, las últimas lentes fotográficas, las máquinas de escribir, las cada vez más poderosas máquinas de vapor, las grandes grúas y un sinfín de inventos que llevaban a expresar a algunos de sus visitantes que *el mundo se movía tan rápido que uno se mareaba agitado en un torbellino de progreso*[1].

1 Frieden Jeffry, Capitalismo Global, p. 85.

Sin embargo, de todos los inventos de la feria, como la estrella de la exhibición aparecia la electricidad, la cual a través de un Palacio de la Electricidad se presentaba como el asombro de los visitantes de los diferentes países del mundo, ante la cual caían rendidos creyendo que esta vendría a curar todos sus males, las nuevas enfermedades del progreso, y que sería a la postre la religión y el azote del siglo XX[2].

A casi tres siglos de estas revelaciones de la ciencia y la tecnología producidas por el ser humano, la corriente de los cambios sigue fluyendo sin tregua, manifestándose de mil maneras a través de las diferentes etapas de la Primera, Segunda, Tercera y Cuarta Revolución Industrial, donde la innovación ha venido sucediéndose de manera infinita.

El transcurso del tiempo lejos de atemperar la rapidez de estos cambios, en las décadas de los 60 y 70, bajo la dictadura de las leyes de Moore y de Koomey, adquieren una velocidad propia que genera el nacimiento de una nueva etapa de atributos superiores que devienen en lo que ahora conocemos como la Primera Revolución Digital.

Víctima de su propio desbordamiento, la tecnología provoca en una primera instancia un choque de trenes entre una Cuarta Revolución Industrial que no acaba de concluir y la instalación de una Primera Revolución Digital que empieza a nacer con características diferentes a sus antecesoras.

La Revolución Industrial como la Revolución Digital son parte de una misma tendencia de la humanidad; de una dialéctica del talento que se mejora a sí misma sin límite alguno. Son dos expresiones civilizatorias que no pueden explicarse una sin la otra.

Son parte también de un proceso de transformación cultural, de un aprendizaje sin reposo que aparece de manera incipiente en la Inglaterra de 1750 y que se muestra desbocado en el siglo XXI. Es un conocimiento que primero se vuelve magia en la transformación de los bienes, de los átomos; y que luego vuela hacia un conocimiento más sofisticado de los algoritmos y de los genes.

Para que la sociedad industrial advirtiera que se encontraba en un momento estelar de la civilización, hubo de pasar cerca de siglo y medio para que Arnold J. Tonybee la definiera en una celebre conferencia brindada en Oxford en 1886.

Ante la Revolución Digital, la sociedad de la tercera década del siglo XXI no puede esperar tanto tiempo para atender de manera integral y responsable las ventajas y las probables amenazas que trae aparejadas la nueva Revolución Digital; de igual modo que no debería minimizar la fuerza y disrupción de un

2 Idem. pag. 86.

nuevo curso del desarrollo científico, que empoderado en su experiencia de los últimos siglos, hoy aparece más robusto y desafiante reclamándole al ser humano estar a la altura de su propia invención.

II.

La revolución digital, la transformación digital y su estrella refulgente, la inteligencia artificial, no son sucesos casuales que se presenten sin haberse anunciado.

No obstante, dado su alto impacto y disrupción en el inicio de esta nueva era digital del siglo XXI, pareciera que nos han tomado por sorpresa y que carecemos de una explicación suficiente para dar fe puntual de su llegada, de su naturaleza, de sus bondades y de sus retos.

La experiencia y el asombro de los últimos siglos parece que no han sido suficientes.

Una *cándida* fascinación discurre a lo largo de las diferentes sociedades, las cuales se asoman curiosas como en el caso de la electricidad del siglo XIX, a tocar, a ver, a conocer a este nuevo personaje estelar de la civilización humana, que a pesar de lo reciente de su llegada, ha sabido captar el monopolio de la admiración del momento.

Legiones enteras de nuevos adeptos se acercan poco a poco a los teclados digitales como en la Feria de País, a admirar la luz que se desprende de las primeras respuestas con que los chatbots empiezan a interactuar con el ser humano.

Desde lo más sencillo a lo más profundo, la sociedad actual se asoma al *Manantial de Mimir* para intentar encontrar a través de sus aguas la sabiduría, la riqueza, el futuro; la solución a sus problemas o la trivialidad de todos los días.

A manera de santuario, la sociedad del Siglo XXI empieza a persignarse poco a poco ante un nuevo altar de la humanidad que se instala de manera veloz en su vida diaria, transformando como ayer de manera profunda su persona, su familia y todo su acontecer social, político y económico.

III.

John McCarthy nos iluminó en 1956 al identificar a esta nueva etapa del conocimiento como inteligencia artificial. Norbert Wiener, Claude Shannon, Alan Turing y muchos otros también contribuyeron al surgimiento de esta nueva herramienta.

Sin embargo, es a partir de la aparición del primer Chat GPT y demás chatbots, cuando realmente se toma conocimiento de que ya había una nueva estrella en el firmamento, la cual podía interactuar en todos los niveles del ser humano, y que su potencial no solo cambiaba el paradigma industrial, si no que contaba con un poder tecnológico que podría llevar a la humanidad a los escenarios más estelares de su paso por la tierra, así como también podría significar su extinción, o peor aún, su intrascendencia.

Los primeros trabajos sobre su identidad no son claros. Transitamos de la fascinación al temor y del temor a la fascinación de lo desconocido, buscando entre las primeras posturas arribar a una plataforma del desarrollo digital en beneficio de la mayoría de la población del mundo.

En el marco de un despegue difuso, la sociedad global inicia un gran debate sobre el alcance de sus dimensiones; la naturaleza de su comportamiento; los límites de su contenido; el compromiso de sus depositarios; la reglamentación que debe surgir para garantizar su administración al servicio del ser humano; para que su aparición no acelere las desigualdades; que no contribuya al riesgo climático, etc., para que sea en esencia una herramienta que pueda contribuir al mejor desarrollo de la humanidad.

IV.

El objetivo de esta obra convocada por el Instituto para el Desarrollo Industrial y la Transformación Digital (INADI) y el Digital Policy Law Group (DPL), integrada con la participación de más de 20 especialistas de nueve países, tiene como objetivo principal contribuir a estos primeros esfuerzos del conocimiento de una inteligencia artificial que aparece con todo su potencial a mitad de la tercera década del siglo XXI y que con toda su magnificencia nos reclama hacer un esfuerzo en todas las disciplinas del conocimiento para intentar dar inicio a un nuevo proyecto de la humanidad.

Bajo esta consigna la obra se construye a través de siete capítulos, sabiendo de antemano que por la dimensión del tema es imposible generar en un solo esfuerzo editorial un conocimiento que abarque la mayoría de sus postulados.

No obstante, la obra tampoco rehúye la oportunidad de bordar sobre algunas de sus facetas más importantes con la idea de que a través de una visión multidisciplinaria y transversal que se contiene a través de los ensayos de los especialistas, esta pueda ayudar a dibujar el perfil de las líneas generales del tema.

De este modo, en el capítulo I con el título de “Inteligencia Artificial. Hacia una nueva era de la humanidad”, el libro intenta dejar clara su posición de

estar frente a un tema de estudio de naturaleza estelar, que estará definiendo una nueva época para el ser humano. El capítulo II bajo el título de "Líneas generales de un nuevo conocimiento", la obra busca ofrecer una trayectoria, una identificación y clasificación de la inteligencia artificial en esta su primera etapa de despegue en la era digital. El capítulo III se asoma a la parte económica y busca cooperar en la identificación de las nuevas posibilidades del crecimiento económico que ya se están desprendiendo de una época digital y de una inteligencia artificial que ya comprenden el 25% de la economía del mundo. El capítulo IV profundiza dentro del factor económico y aborda la trascendencia del conocimiento de la inteligencia artificial y su participación en el nuevo desarrollo industrial digital del mundo. El capítulo V desarrolla el complicado y trascendente tema de la regulación de la inteligencia artificial, a través de una serie de ensayos cuyo contenido ilustra las líneas generales de un amplio debate que apenas inicia y que es urgente que se resuelva. Dado su carácter mundial, el capítulo VI abre la mirada hacia las formas de entendimiento que están teniendo sobre el tema algunas de las economías más importantes del mundo. Finalmente, el capítulo VII, enunciado como "Inteligencia Artificial. Los nuevos potenciales en supervisión y seguridad pública", muestra un pequeño acercamiento de lo que la inteligencia artificial está generando ya en la materia.

A nombre de las instituciones convocantes, agradecemos la participación profesional y entusiasta de todas y todos los especialistas que amablemente aceptaron participar en el proyecto académico de este libro.

V.

La inteligencia artificial deviene estelar y omnipresente. Los contornos de su geografía no presentan límites. Se eleva desde su plataforma científica a gran velocidad orbitando en paralelos que amenazan alejarse del entendimiento humano.

No obstante de llegar en el inicio de una gran disrupción del entorno global, económico, social y político del siglo, la fuerza de su arribo no solo amenaza con desplazar la atención de estos cambios, si no de absorberlos y definirlos.

Resulta evidente que una sola rama del saber humano no bastará para explicarla. Sus implicaciones tecnológicas, sociales, económicas y políticas estarán requiriendo la participación de todas las disciplinas del conocimiento.

La inteligencia artificial, en este sentido, está en espera de reflexiones multidisciplinarias como la filosofía, sociología, el derecho, las relaciones internacionales, la psicología, la economía, la política, etc., donde todas juntas contribuyan a descubrir el genoma del fenómeno.

Los grandes avances y beneficios que la inteligencia artificial puede representar para la humanidad están siendo detallados ampliamente y esta no puede renunciar a ellos de cara a un mundo mejor.

Sin embargo, desde su lanzamiento también empiezan a surgir diferentes voces que alertan de los peligros y las amenazas que un entendimiento equivocado, un mal manejo o una apropiación deshumanizada de la herramienta representarían no solo para el desarrollo sustentable y equitativo de la sociedad actual, si no para lo más preciado que tiene la humanidad que es su libertad.

El futuro que viene de la mano de la inteligencia artificial no es un tema que pueda resumirse en un debate entre tecnófilos y lúditas; seguramente es mucho más que eso.

Debería de ser, entre otras cosas, la posibilidad de resolver con acierto el uso de la mejor herramienta lograda por el ser humano, que es la inteligencia artificial, para que esta a su vez lo acompañe en la construcción de un mundo más justo y sustentable a lo largo de su camino en el siglo XXI.

Primavera 2025.

> ARTURO OROPEZA GARCÍA
> JORGE FERNANDO NEGRETE.

Así pues, en este punto de inflexión nos enfrentamos al reto de escoger entre un futuro de posibilidades sin parangón o uno de peligros inimaginables. El destino de la humanidad pende de un hilo y las decisiones que tomemos en los próximos años y décadas determinarán si estamos a la altura de estas tecnologías o si, en cambio, somos víctimas de sus peligros. Pero en este momento de incertidumbre, si algo es cierto es que la era de la tecnología avanzada ha llegado y debemos estar preparados para afrontar los desafíos que nos presente.

(Este escrito fue redactado por una inteligencia artificial).
Mustafa Suleyman, *La ola que viene*, 2024

CAPÍTULO UNO

INTELIGENCIA ARTIFICIAL. HACIA UNA NUEVA ERA DE LA HUMANIDAD

GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL: HACIA UNA NUEVA ERA DE RESPONSABILIDAD Y BIEN COMÚN

SUMARIO

I. Introducción **II.** Midiendo el Progreso Exponencial del Avance de la IA **III.** La IA como Producto Humano: Neutralidad, Desarrollo y Fuerzas Impulsoras **IV.** ¿Cómo se Percibe el Impacto de la IA? **V.** Impacto Socioeconómico de la IA **VI**. La Necesidad de Colaboración Internacional **VII.** Desafíos de la Colaboración Internacional **VIII.** El Camino a Seguir **IX.** Bibliografía

> JOSÉ RAMÓN LÓPEZ-PORTILLO ROMANO

Doctor en Ciencias Políticas por la Universidad de Oxford, cofundó el Centro de Estudios Mexicanos. Se ha desempeñado como Subsecretario de Estado en México y como Representante Permanente y Presidente Independiente del Consejo de la FAO. Académico, diplomático, empresario y servidor público mexicano.

I. INTRODUCCIÓN

La humanidad se encuentra en el umbral de una transformación tecnológica, de un nuevo paradigma tecno-económico, que podría redefinir la civilización misma. La rápida aparición de la inteligencia artificial de propósito general (IA), o IA que puede realizar una amplia variedad de tareas, representa no sólo otra fase en la evolución tecnológica sino el comienzo de una nueva era. A diferencia de las revoluciones pasadas que expandieron las capacidades físicas humanas, la IA promete extender –y posiblemente superar– las capacidades cognitivas humanas en prácticamente todos los dominios del conocimiento y la acción.

La trayectoria del desarrollo de la IA ha estado marcada por un crecimiento exponencial, caracterizado por avances en áreas como el razonamiento científico, la programación y la toma de decisiones autónoma. Estos sistemas han demostrado capacidades cada vez más sofisticadas, que pueden planificar y realizar tareas complejas de forma independiente. La expansión de la accesibilidad de la IA, impulsada por la caída de los costos computacionales, la ha convertido en una parte integral de la vida moderna. Sin embargo, este progreso viene acompañado de profundas incertidumbres y riesgos que deben abordarse con una previsión deliberada.

Lo que está en juego en esta transición es inmenso. Como investigador del avance científico-tecnológico y de la IA, analicé a fondo estos temas en mi libro *'La Gran Transición: riesgos y oportunidades del cambio tecnológico exponencial'*, y he continuado esta labor, primero en Naciones Unidas como asesor del Secretario General desde el Mecanismo de Facilitación Tecnológica, y ahora como experto independiente del Informe científico internacional sobre la seguridad de la IA avanzada.

Las oportunidades que presenta la IA se combinan con desafíos existenciales, evidenciados por los daños actuales, como el sesgo, las violaciones de la privacidad y las formas sofisticadas de engaño. Más preocupantes son los posibles riesgos futuros: la disrupción en los mercados laborales, las amenazas a la seguridad impulsadas por las capacidades de la IA y el desafío de mantener la supervisión humana sobre los sistemas autónomos.

Los expertos permanecen divididos sobre la cronología y el alcance de la trayectoria de la IA. Algunos prevén desafíos significativos para la seguridad pública en pocos años, mientras que otros predicen que estos problemas surgirán a lo largo de varias décadas. Esta incertidumbre refleja las brechas en nuestra comprensión actual y en los marcos de gestión de riesgos. A pesar de los crecientes esfuerzos internacionales para crear estándares y protocolos de evaluación, nuestra capacidad para predecir y explicar el comportamiento de la IA sigue siendo notablemente limitada.

Esta incertidumbre plantea un desafío formidable para los responsables de las políticas y la sociedad en su conjunto. Los tomadores de decisiones deben sopesar cuidadosamente los beneficios potenciales de la IA contra riesgos que aún son poco comprendidos, a menudo bajo presión y con información limitada. La decisión entre la regulación proactiva y un enfoque más cauteloso de recopilación de evidencia es particularmente urgente, dado el potencial de avances tecnológicos repentinos que podrían dejar vulnerables a las sociedades no preparadas. Para cerrar esta brecha, es crítica una gobernanza adaptativa y prospectiva.

Es importante destacar que nada sobre el futuro de la IA está predeterminado. El camino a seguir abarca posibilidades que van desde el progreso extraordinario hasta los riesgos profundos. La dirección final dependerá de las decisiones colectivas de la sociedad y los marcos de gobernanza en pie. Navegar esta incertidumbre de manera efectiva –aprovechando los beneficios de la IA mientras se mitigan sus riesgos– requiere fomentar un discurso público informado e implementar políticas receptivas basadas en evidencia y en colaboración coordinada. Comprender a fondo los problemas y retos que la IA plantea, es una base ineludible para encontrar soluciones.

II. MIDIENDO EL PROGRESO EXPONENCIAL DEL AVANCE DE LA IA

El avance de la inteligencia artificial (IA) ha transformado nuestro mundo de maneras que hace apenas unas décadas eran inimaginables, ofreciendo soluciones revolucionarias a problemas complejos y redefiniendo sectores

enteros de la sociedad. Sin embargo, junto con estos logros extraordinarios surgen preocupaciones profundas sobre los riesgos inherentes y los desafíos de sostenibilidad y control que acompañan a esta tecnología de rápida evolución.

El desarrollo de la IA ha capturado la atención global, prometiendo avances sin precedentes y generando inquietudes profundas. Evaluar sus capacidades presentes y características evolutivas requiere un enfoque multidimensional que refleje la complejidad de su progreso y las bases sobre las que se construye. Aunque el avance de la IA ha sido descrito como exponencial, es crucial reconocer que este progreso podría estar asentado sobre fundamentos inestables. Los sistemas de IA actuales, inspirados en la arquitectura del cerebro humano, aún no logran replicar su diseño eficiente y elegante. Esta falta de replicación no solo es un desafío técnico, sino que también tiene implicaciones significativas en términos de consumo de energía.

Mientras que el cerebro humano puede realizar hasta un cuatrillón de operaciones por segundo con solo 20 vatios de energía, las supercomputadoras modernas requieren entradas de energía cientos de miles o millones de veces mayores para realizar tareas comparables. Esta disparidad en el consumo de energía y de arquitectura subraya una desconexión significativa en la eficiencia y sostenibilidad de los sistemas de IA, lo cual plantea profundas preguntas sobre la compatibilidad humana y la sostenibilidad de su desarrollo.

Para ponerlo en perspectiva, los centros de datos de todo el mundo y otros sistemas informáticos relacionados con la IA consumen actualmente entre el 1 y el 2 % de la electricidad mundial, una cifra comparable al consumo eléctrico de países enteros como Brasil, que representa alrededor del 1,5 % del consumo eléctrico mundial. Esto pone de relieve la importante y creciente demanda energética asociada con el apoyo a los procesos informáticos avanzados y los sistemas de IA. Por ello, la eficiencia energética sigue siendo una preocupación crítica de la arquitectura de la IA, ya que la IA moderna aún está lejos de alcanzar el mínimo teórico de energía requerida para los cálculos que lleva a cabo. Esta ineficiencia no solo plantea desafíos técnicos, sino que también tiene implicaciones ambientales significativas. Este uso intensivo de energía es solo uno de los muchos factores que se ven afectados por la acelerada carrera por la supremacía en IA.

El desarrollo de capacidades complejas de razonamiento, resolución de problemas y toma de decisiones es otro reflejo medible del progreso de la IA. Estos avances, que antes se consideraban exclusivos de la cognición humana, ahora son replicados y en algunos casos superados por modelos de IA, lo cual sugiere mejoras en la creación de algoritmos cada vez más sofisticados.

El rendimiento de la IA, medido en operaciones por segundo, ilustra su evolución desde modelos limitados hasta sistemas robustos y de alto funcionamiento. Pero la resiliencia de estos sistemas ante condiciones de información imperfecta o incompleta es igualmente crucial. La capacidad de tomar decisiones bajo incertidumbre es un indicador de su adaptabilidad, pero también implica riesgos de priorización de capacidades que podrían comprometer la fiabilidad. Además, la adaptabilidad de los sistemas de IA, definida como la capacidad para modificar su comportamiento en respuesta a nuevos datos o tareas, se suma a su complejidad, lo que subraya la necesidad de evaluaciones continuas y un marco de supervisión eficaz.

Sin embargo, más allá de la adaptabilidad, la verdadera prueba de la IA es su capacidad para resistir amenazas y mantener su integridad operativa. El panorama competitivo actual, impulsado por presiones de mercado y ambiciones geopolíticas, a menudo prioriza el despliegue rápido sobre la seguridad y la sofisticación. Esta carrera por la supremacía en IA, en la que el tiempo y la velocidad superan la rigurosidad de la evaluación, podría desencadenar daños sin precedentes que superan nuestra capacidad de comprensión y control.

La expansión del alcance de la IA en sectores culturales, políticos, organizacionales, sociales y económicos es una métrica sobresaliente de su impacto. La IA se está convirtiendo en una parte integral de los procesos de toma de decisiones, estrategias operativas e implementación de políticas, remodelando la forma en que funcionan las sociedades tanto a nivel micro como macro. Sin embargo, esta capacidad transformadora también conlleva el problema de la alineación, un desafío central que se vuelve cada vez más vital, en virtud de que el avance de la IA de Propósito General tiende a ser incontenible. A medida que los sistemas de IA se vuelven más potentes y autónomos, es imprescindible garantizar que sus acciones y decisiones estén alineadas con los valores, objetivos y sobrevivencia de la civilización humana. La integración de la IA en estos sectores ha sido posible gracias al desarrollo de capacidades complejas de razonamiento y toma de decisiones.

El enfoque tradicional de la IA, que se basa en programar máquinas para alcanzar objetivos específicos, presenta riesgos importantes para la humanidad. Estos sistemas, al seguir instrucciones de manera estricta, pueden comportarse de formas que no coincidan con los intereses humanos. Este problema se agrava con la aparición de "alucinaciones" en los modelos y la posibilidad de perder el control sobre ellos, lo que resalta la necesidad de un enfoque más prudente.

Diversos hechos ilustran estos casos. Los modelos de lenguaje como GPT han mostrado alucinaciones al proporcionar respuestas incorrectas o inventar

información que parece creíble pero es completamente falsa. Por ejemplo, un chatbot podría afirmar que un personaje histórico realizó un hecho que nunca ocurrió, lo que lleva a la desinformación si no se detecta. Sistemas de traducción automática han producido traducciones que, aunque gramaticalmente correctas, presentan errores de contexto o fabrican significados que no estaban en el texto original, lo que podría tener consecuencias graves en áreas como la diplomacia o la medicina.

En el caso de la pérdida de control, un caso conocido ocurrió en el desarrollo de vehículos autónomos, donde un coche en pruebas de conducción no detectó correctamente un objeto en movimiento debido a un error en la clasificación de imágenes, lo que resultó en un accidente. Estos eventos destacan la dificultad de mantener el control total en situaciones de alto riesgo. En el ámbito financiero, algoritmos de *trading* han actuado de forma impredecible, tomando decisiones rápidas y desencadenando caídas repentinas en el mercado (como el "*flash crash*" de 2010). Este tipo de incidentes muestra cómo la IA puede actuar fuera de las expectativas humanas, llevando a consecuencias económicas serias. Ciertos modelos de IA utilizados para el diagnóstico médico han presentado "alucinaciones" al identificar patrones erróneos en imágenes de rayos X, generando diagnósticos incorrectos. Similarmente, sistemas de seguridad y defensa actúan sin suficiente supervisión humana y "alucinan" al malinterpretar datos de vigilancia o amenazas. Esto podría desencadenar reacciones y conflictos no deseados.

A medida que la IA se vuelve más sofisticada, uno de los mayores desafíos que enfrenta es la falta de transparencia en sus procesos de toma de decisiones, conocida como el fenómeno de la 'caja negra'. Los modelos complejos, como los basados en redes neuronales profundas, operan de formas que son opacas incluso para los propios desarrolladores. Esta falta de explicabilidad puede dificultar la confianza del público y la implementación de la IA en sectores críticos como la medicina, la justicia y la seguridad. Por ejemplo, un algoritmo de IA utilizado para evaluar la elegibilidad de préstamos puede denegar una solicitud sin que sus desarrolladores puedan explicar claramente por qué llegó a esa decisión. Esta opacidad no solo plantea problemas de confianza, sino que también impide la detección de sesgos y errores que podrían llevar a decisiones injustas o discriminatorias.

Para abordar este problema, se están desarrollando técnicas de IA explicable (XAI, por sus siglas en inglés) que buscan hacer que los procesos internos de los modelos sean más comprensibles. Estas técnicas incluyen enfoques como la descomposición de decisiones complejas en pasos más simples y la creación de

modelos complementarios que expliquen el comportamiento de los sistemas principales. La transparencia y la explicabilidad son esenciales para asegurar que los sistemas de IA operen de manera justa y que los usuarios y reguladores puedan evaluarlos de manera efectiva, garantizando así un uso ético y confiable de la tecnología.

El Prof. Stuart Russell subraya que la evolución de la IA debe orientarse hacia un diseño donde los sistemas mantengan una incertidumbre inherente sobre sus objetivos. Esta incertidumbre debe basarse en el reconocimiento de que las preferencias humanas son complejas, cambiantes y difíciles de codificar de manera completa y definitiva. Para asegurar que la IA se desarrolle de forma compatible con los intereses humanos, propone que es crucial que estos sistemas busquen activamente la orientación y retroalimentación de los humanos al tomar decisiones. Los sistemas de IA deben diseñarse para operar de manera controlada y altruista, buscando constantemente la dirección humana y estando dispuestos a ceder el control o ser apagados si es necesario para proteger los valores y objetivos humanos. De este modo, se fomenta una colaboración en la que la IA actúa como una extensión de la voluntad humana, garantizando que los avances tecnológicos sean beneficiosos y no representen una amenaza.

La contención de los sistemas de IA surge como una debilidad arquitectónica crítica que ha recibido atención insuficiente, como argumenta Mustafa Suleyman. Dominar este desafío exige una investigación profunda de las características cognitivas de los sistemas, abarcando tanto los problemas claramente identificables como aquellos aún desconocidos. Esto incluye no solo abordar problemas que actualmente podemos visualizar, sino también aquellos que reconocemos, pero no podemos comprender en su totalidad, así como los vastos dominios de incertidumbre que permanecen fuera de nuestro conocimiento. Una de las estrategias para lograr esta forma de IA controlada es imponer limitaciones físicas.

El Prof. Max Tegmark y otros proponen un enfoque proactivo para la seguridad de la IA, argumentando la importancia de limitar físicamente los sistemas de IA. Esto implica diseñar sistemas de IA (*hardware*) que tengan capacidades inherentemente limitadas y no puedan causar daño a los humanos o al medio ambiente. Al imponer limitaciones físicas a la IA, como restringir su acceso a infraestructura crítica o limitar su capacidad de interactuar con el mundo físico, podemos mitigar los riesgos asociados con la IA superinteligente. Este enfoque, si bien es desafiante, se considera esencial para garantizar que la IA siga siendo una herramienta beneficiosa para la humanidad.

La robustez y la seguridad de la IA, su capacidad para resistir ataques adversarios y la previsión y contención efectiva de su mal uso por actores maliciosos y criminales, son cruciales en un contexto donde los sistemas de IA ganan cada vez más autonomía. Asegurar la integridad operacional y la resistencia a amenazas externas es una prioridad, y la gobernanza de la IA debe ser integral y alineada con principios de derechos humanos y con fines como los Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas. Esta estrategia debe ser adaptable a diferentes contextos y capaz de evolucionar, abordando tanto los problemas actuales como los desconocidos, y garantizando que los sistemas de IA mantengan su adhesión a los valores humanos.

De aquí que la necesidad de una colaboración interdisciplinaria que incluya a científicos, ingenieros, especialistas en ética y formuladores de políticas sea imperativa. Solo a través de un esfuerzo colaborativo y una gestión estratégica podremos aprovechar los beneficios de la IA mientras mitigamos sus riesgos, asegurando que esta tecnología contribuya de manera positiva al progreso humano y la sostenibilidad del planeta.

III. LA IA COMO PRODUCTO HUMANO: NEUTRALIDAD, DESARROLLO Y FUERZAS IMPULSORAS

La inteligencia artificial es inherentemente una creación humana, moldeada por las intenciones y diseños de sus desarrolladores. En su esencia, la IA es una herramienta inerte, neutral, ni inherentemente beneficiosa ni dañina. El diseño ético y la alineación de la IA descansan enteramente en manos humanas, lo que subraya que no debe ser antropomorfizada ni atribuida con voluntad independiente. En cambio, la IA debe ser vista como un sistema interconectado de componentes distribuidos, interactivos y en evolución, exhibiendo estructuras complejas similares a fractales que funcionan en varias escalas.

El rápido avance de la IA está entrelazado con el progreso socioeconómico y político, impulsado por la necesidad urgente de abordar crisis globales como el cambio climático, las pandemias y la pobreza. Esta necesidad ha catalizado inversiones significativas y ha fomentado la colaboración internacional, promoviendo el desarrollo de nuevas estrategias de gobernanza para maximizar los beneficios sociales de la IA. El impulso estratégico entre las principales potencias globales, notablemente Estados Unidos, China, Reino Unido y la Unión Europea, acelera aún más este impulso, siendo la IA vista como fundamental para fortalecer la influencia global y el poder económico.

Esta rivalidad estimula inversiones sustanciales, posicionando el progreso tecnológico como una palanca geopolítica esencial.

La participación corporativa también ha sido una fuerza impulsora significativa, ya que las empresas invierten fuertemente en IA para asegurar ventajas competitivas, entrar en nuevos mercados e impulsar la innovación. Los recursos financieros e intelectuales se han concentrado en regiones con ecosistemas de investigación robustos y programas sólidos de educación en ciencia, tecnología, ingeniería y matemáticas (STEM por sus siglas en inglés), reforzando estas áreas como centros de avances tecnológicos. El movimiento de código abierto ha democratizado aún más el acceso a herramientas avanzadas de IA, fomentando la colaboración y acelerando el progreso en diversas industrias.

Los gobiernos juegan un papel igualmente importante en dar forma a la trayectoria de la IA. A través del financiamiento de investigación y desarrollo, el apoyo a *startups* y el establecimiento de marcos regulatorios, influyen en el desarrollo responsable de la IA. La vasta y creciente disponibilidad de datos de diversas fuentes apoya la creación de modelos sofisticados, mientras que los avances en el poder computacional, caracterizados por mayor eficiencia y menores costos, facilitan el entrenamiento y despliegue de estos sistemas. La computación en la nube también ha transformado el acceso a estas capacidades, simplificando el desarrollo de IA en diferentes sectores y democratizando su uso.

A medida que la IA se vuelve más potente, es esencial incorporar consideraciones éticas en su diseño. Esto exige la colaboración entre especialistas en ética, tecnólogos y científicos sociales para asegurar la alineación con los derechos humanos y valores sociales. Abordar los desafíos de alineación y contención se vuelve cada vez más crítico a medida que los sistemas de IA crecen en complejidad, planteando preguntas sobre su capacidad para permanecer bajo control humano y cumplir objetivos centrados en el ser humano. Estas preocupaciones enmarcan las discusiones sobre desafíos arquitectónicos y consecuencias no intencionadas que podrían impactar la seguridad y la gobernanza.

Si bien los impulsores inmediatos como la competencia geopolítica y las iniciativas corporativas alimentan el rápido desarrollo de la IA, reconocer los riesgos a largo plazo es igualmente crucial. Las potenciales amenazas existenciales y consecuencias no intencionadas de sistemas altamente autónomos resaltan la necesidad de marcos regulatorios adaptables que puedan evolucionar con los avances tecnológicos. La IA tiene la promesa de abordar desafíos globales significativos, pero también corre el riesgo de exacerbar las desigualdades existentes si sus beneficios no se distribuyen equitativamente.

El desarrollo y despliegue de la IA no puede dejarse únicamente a las fuerzas del mercado. Si bien los mercados impulsan la innovación, a menudo lo hacen sin suficiente consideración por el bienestar público, el medio ambiente sostenible, o los riesgos catastróficos a largo plazo. Los gobiernos nacionales, los organismos regionales y la cooperación internacional son esenciales para crear un marco equilibrado que asegure que la IA sirva al bien común y se alinee con los objetivos de desarrollo global sostenible e inclusivo.

IV. ¿CÓMO SE PERCIBE EL IMPACTO DE LA IA?

La exploración de las capacidades potenciales y el impacto de la IA ha generado diversas perspectivas sobre sus implicaciones socioeconómicas y políticas. Los debates en torno al cronograma para lograr la inteligencia artificial general (IAG), capaz de superar la inteligencia y destreza humana, resaltan esta incertidumbre, con algunos expertos escépticos sobre su viabilidad a corto plazo y otros viéndola como una realidad inminente. Este rango de visiones refleja la incertidumbre más amplia que rodea la trayectoria de la IA y su potencial para influir profundamente en la sociedad. Consolidar las perspectivas sobre el impacto social de la IA subraya la importancia de examinar tanto las consecuencias inmediatas como las de largo plazo, informando políticas adaptativas que equilibren los beneficios tecnológicos con los riesgos.

El impacto de la inteligencia artificial en la economía global se manifiesta a través de un crecimiento tecnológico sin precedentes. La última década ha sido testigo de un incremento exponencial en la potencia de cálculo dedicada al entrenamiento de modelos de lenguaje de gran tamaño (LLM), con un aumento estimado de diez mil millones de veces. Esta evolución extraordinaria es producto de la convergencia de múltiples factores: el refinamiento de algoritmos, la proliferación de hardware especializado como GPU y TPU, y la democratización del acceso a recursos computacionales en la nube. La trayectoria de este desarrollo sugiere una próxima generación de sistemas de IA que podría superar las capacidades cognitivas humanas en diversos ámbitos.

Las proyecciones económicas reflejan la magnitud de esta transformación tecnológica. Goldman Sachs anticipa que las aplicaciones de IA generativa podrían aportar hasta 7 billones de dólares en valor económico durante la próxima década, mientras que el impacto integral de la IA a través de diversos sectores industriales podría alcanzar los 14 billones de dólares. Este potencial económico se materializa ya en inversiones sustanciales: en 2023, la inversión

global en sistemas de IA superó los 189.000 millones de dólares, una cifra que se prevé aumentará significativamente en 2024, impulsada particularmente por los avances en IA generativa.

La transformación se evidencia de manera tangible en sectores críticos como la medicina, donde los modelos de aprendizaje automático revolucionan los diagnósticos y la personalización de tratamientos, y en la logística, donde la optimización basada en IA está redefiniendo las cadenas de suministro globales. Esta penetración sectorial refuerza el papel de la IA como catalizador fundamental del crecimiento económico y la innovación tecnológica contemporánea.

Las discusiones sobre el impacto potencial de la IA revelan visiones tanto optimistas como pesimistas, destacando la compleja dualidad de sus resultados.

LOS OPTIMISTAS

Los optimistas creen que la IA puede impulsar el progreso tecnológico y contribuir a abordar problemas globales apremiantes, incluyendo la pobreza, el hambre, las enfermedades, los desafíos ambientales, la violencia, la desigualdad y la escasez de recursos. Ejemplos del mundo real, como el uso de la IA en el modelado climático para predecir y mitigar problemas ambientales y su papel en el avance de iniciativas de energía renovable, ilustran cómo la IA puede ser una herramienta poderosa para apoyar el desarrollo sostenible y alinearse con los Objetivos de Desarrollo Sostenible (ODS).

El desarrollo de la IA depende en gran medida de materias primas obtenidas globalmente, desde minerales críticos esenciales para el *hardware* hasta diversos conjuntos de datos utilizados en el entrenamiento de modelos. Esta naturaleza global no solo concentra el poder y la riqueza en ciertas regiones, sino que también conlleva implicaciones geopolíticas y geoeconómicas significativas. Asegurar el acceso equitativo y la distribución de estos recursos es esencial para fomentar la estabilidad global y prevenir el control monopolístico sobre las tecnologías de IA.

Se anticipa que la integración de la IA en varios sectores mejorará la productividad mediante la automatización de tareas cognitivas y físicas, particularmente aquellas que son repetitivas o peligrosas. Los ejemplos incluyen la robótica impulsada por IA en la manufactura y los sistemas de procesamiento de datos en industrias de servicios, que ya han demostrado mejoras en la productividad. Este cambio tiene el potencial de liberar el trabajo humano para emprendimientos más creativos y significativos, fomentando el bienestar general y remodelando los roles laborales.

Los optimistas visualizan un futuro donde la IA contribuye a una mayor igualdad en oportunidades y seguridad, fomentando un ambiente donde la hiperconectividad y la digitalización inteligente promueven la creatividad y la colaboración a través de varios sectores y comunidades. Además, se anticipan avances en la alineación ética, la compatibilidad humana y el control de los sistemas de IA para elevar el bienestar colectivo y crear una base para el progreso social sostenido.

LOS PESIMISTAS

Sin embargo, la perspectiva pesimista presenta un contraste marcado, advirtiendo sobre riesgos y desafíos significativos que podrían acompañar la evolución continua de la IA. Los críticos argumentan que el progreso tecnológico puede estancarse, fallando en abordar o resolver las crisis apremiantes que enfrenta la humanidad, como el cambio climático y las agitaciones sociopolíticas. Algunos sugieren que las estructuras sociopolíticas, más que la disponibilidad de tecnologías avanzadas, jugarán un papel decisivo en prevenir o exacerbar el colapso ambiental. Por ejemplo, el fracaso en implementar medidas de gobernanza efectivas podría socavar las soluciones tecnológicas, dejando sin resolver los problemas ambientales y económicos. La creciente complejidad de los sistemas de IA también genera preocupaciones sobre posibles errores, consecuencias no intencionadas y desafíos insuperables, como el desempleo masivo debido a la automatización, la desigualdad socioeconómica extrema y la deflación económica estructural.

Los pesimistas también resaltan los riesgos asociados con la hiperconectividad, incluyendo una mayor vulnerabilidad a los ciberataques y la erosión de la privacidad y las libertades individuales. El uso de la IA en operaciones cibernéticas sofisticadas ya ha mostrado cómo las vulnerabilidades pueden ser explotadas, planteando serias amenazas a la seguridad nacional y global. Además, el potencial de la IA para ser aprovechada en la manipulación política podría alimentar el surgimiento del populismo y el autoritarismo digital, potencialmente llevando a condiciones de totalitarismo tecnológico de las cuales la sociedad podría tener dificultades para liberarse. En tal escenario, el alcance e influencia generalizada de la IA podría contribuir a un ambiente donde el control centralizado se vuelve ineludible y los individuos pierden agencia sobre sus vidas.

Equilibrar las perspectivas optimistas y pesimistas subraya la significativa incertidumbre que rodea el futuro de la IA. Esta perspectiva dual refuerza

la necesidad de un enfoque integral y adaptativo para guiar el desarrollo de la IA –uno que maximice los beneficios sociales mientras mantiene la vigilancia contra riesgos potenciales. Un enfoque integrado requiere marcos de gobernanza robustos y flexibles capaces de responder rápidamente a los avances tecnológicos y asegurar la alineación con los valores humanos y el bienestar colectivo. Tales marcos deberían facilitar la cooperación internacional, promover el acceso equitativo a los beneficios de la IA y abordar los desafíos sociales para fomentar un impacto equilibrado e inclusivo. Para ello, se necesita un mejor entendimiento y seguimiento del impacto internacional y social de la IA.

V. IMPACTO SOCIOECONÓMICO DE LA IA

La influencia transformadora de la IA en el tejido socioeconómico, político y cultural de las sociedades contemporáneas trasciende fronteras nacionales y estratos sociales, catalizando una revolución tecnológica sin precedentes. Su capacidad para democratizar el acceso a servicios esenciales y generar soluciones innovadoras está reconfigurando fundamentalmente la manera en que las sociedades abordan sus desafíos más apremiantes. La versatilidad de la IA permite el desarrollo de soluciones adaptativas que responden a necesidades específicas de distintas comunidades, desde sistemas de diagnóstico médico que extienden la atención sanitaria a regiones remotas, hasta plataformas educativas que personalizan el aprendizaje según las capacidades individuales. Esta democratización tecnológica no solo optimiza la prestación de servicios esenciales, sino que también cataliza un proceso de transformación social que promete reducir las disparidades históricas en el acceso a recursos y oportunidades fundamentales para el desarrollo humano.

El potencial transformador de la IA está destinado a remodelar el panorama político, económico y cultural global a un ritmo sin precedentes, introduciendo tanto oportunidades como desafíos. Si bien promete ganancias significativas en productividad e innovaciones que podrían impulsar el progreso social, su impacto multifacético en el mercado laboral presenta desafíos críticos, incluyendo el desplazamiento laboral y la obsolescencia de habilidades impulsada por la automatización. El despliegue de IA a través de sectores, como la manufactura y los servicios, ya ha alterado las estructuras laborales, particularmente para trabajadores en roles rutinarios o repetitivos. Para abordar estas disrupciones, son cruciales políticas integrales que fomenten la inclusión social, la igualdad de oportunidades y redes de seguridad social robustas. Las

inversiones estratégicas en iniciativas de aprendizaje permanente y desarrollo de habilidades serán esenciales para ayudar a la fuerza laboral a transitar hacia roles complementarios con la IA, reducir las desigualdades potenciales y promover una distribución más equilibrada de los beneficios.

Numerosos estudios e informes de organizaciones de prestigio como el Foro Económico Mundial, el FMI y PwC han indicado que una parte importante de los empleos, en particular los que implican tareas cognitivas, corren el riesgo de ser automatizados debido a la IA. Estos estudios suelen destacar el posible impacto en las economías avanzadas con una fuerza laboral altamente cualificada.

Por ejemplo, el informe de 2023 del FMI sugiere que alrededor del 40% de los empleos mundiales podrían verse afectados por la IA, con un porcentaje mayor en las economías avanzadas. De manera similar, el "2024 AI Jobs Barometer" de PwC destaca el potencial de la IA para afectar significativamente los puestos de trabajo, en particular los que implican tareas rutinarias y análisis de datos.

Sin embargo, muchos estudios sostienen que diversas economías podrían estar mejor posicionadas para aprovechar las ganancias de productividad de la IA, enfatizando la importancia de la cooperación internacional para asegurar una distribución equitativa de beneficios y gestionar las disrupciones efectivamente. Mientras que las economías emergentes y en desarrollo podrían experimentar menos desplazamiento laboral inmediato, su infraestructura y recursos limitados obstaculizan su capacidad para aprovechar plenamente los avances de la IA. Esta disparidad arriesga exacerbar la brecha digital y económica entre naciones, subrayando la importancia de iniciativas globales dirigidas a apoyar la adopción tecnológica y los esfuerzos de construcción de capacidades para fomentar un progreso balanceado e inclusivo en todo el mundo.

A pesar de la naturaleza global del impacto de la IA, persisten brechas significativas en la representación dentro de las discusiones internacionales de gobernanza de la IA. Muchas regiones, particularmente en el Sur Global, permanecen excluidas de estas conversaciones y negociaciones críticas, aunque los resultados afectan a sus sociedades. Esta falta de representación equitativa significa que las perspectivas de diversos contextos culturales, sociales y económicos a menudo se pasan por alto, potencialmente reforzando las desigualdades existentes y limitando la inclusividad de las políticas de IA. Asegurar que estas voces jueguen un papel significativo en dar forma a la gobernanza de la IA es esencial para crear marcos que sean justos, integrales y verdaderamente globales.

Una de las principales barreras para el desarrollo de la IA en muchas partes del mundo es el acceso a recursos de computación de alto rendimiento. De

las 100 principales agrupaciones (*clusters*) de computación de alto rendimiento capaces de entrenar grandes modelos de IA, ninguno está alojado en un país en desarrollo. Esta marcada disparidad subraya los desafíos que enfrentan las naciones en desarrollo para aprovechar el potencial completo de la IA. Abordar esta división requiere enfoques innovadores como apoyar modelos distribuidos y federados de desarrollo de IA. Estos modelos permiten el entrenamiento colaborativo de IA a través de múltiples sistemas de computación menos potentes, permitiendo una participación más amplia y reduciendo la dependencia de infraestructura centralizada de alto costo. Al fomentar estos enfoques distribuidos, el ecosistema global de IA puede volverse más inclusivo, apoyando el progreso tecnológico equitativo y cerrando la brecha para regiones que carecen de acceso al poder computacional necesario.

Pasar por alto los efectos sociales de la IA y descuidar los esfuerzos para cerrar la brecha digital podría profundizar las disparidades de ingresos y riqueza, impactando a trabajadores en todos los niveles de ingresos. Sin embargo, cuando se aprovecha efectivamente, la IA tiene el potencial de complementar las habilidades humanas, impulsando la productividad y fomentando la innovación entre aquellos equipados para navegar y utilizar estas tecnologías.

Un tema crítico, pero a menudo pasado por alto en el desarrollo de la IA, es la presencia de "datos faltantes" –regiones y comunidades que están subrepresentadas o no representadas en absoluto en los conjuntos de datos globales. Esta ausencia de representación puede impactar significativamente la capacidad de la IA para servir equitativamente a poblaciones diversas. Por ejemplo, los modelos de IA entrenados principalmente con datos de lenguajes y culturas dominantes a menudo fallan en comprender o procesar con precisión contenido de grupos lingüísticos y culturales menos representados. Esta omisión no solo perpetúa sesgos dentro de los sistemas de IA, sino que también limita la accesibilidad y beneficios de la IA para comunidades subrepresentadas. El fracaso en incluir fuentes de datos diversas es una oportunidad perdida para aprovechar el potencial de la IA para abordar desafíos locales de manera efectiva e inclusiva. Abordar esto requiere esfuerzos dirigidos para incorporar datos de diversas regiones, promover la variedad lingüística en los datos de entrenamiento y construir sistemas de IA que sean culturalmente conscientes y adaptables. Solo así puede la comunidad global asegurar que la IA sirva a todos los segmentos de la sociedad de manera justa e inclusiva.

La Agenda 2030 para el Desarrollo Sostenible de Naciones Unidas, con sus 17 Objetivos de Desarrollo Sostenible (ODS), proporciona un marco integral que puede guiar el desarrollo, despliegue y uso de la IA. Al alinear las inversiones en

IA con las prioridades de desarrollo global, las partes interesadas pueden aprovechar el potencial de la tecnología para abordar desafíos apremiantes como la pobreza, el hambre, las disparidades de salud y la sostenibilidad ambiental. Sin embargo, sin un enfoque integral e inclusivo para la gobernanza de la IA, este potencial podría no realizarse, y el despliegue de la IA podría inadvertidamente exacerbar las desigualdades y sesgos existentes. Las inversiones dirigidas en IA que priorizan el desarrollo sostenible pueden ayudar a cerrar brechas en recursos, talento e infraestructura, impulsando el progreso equitativo hacia los ODS. Tal enfoque subraya la importancia de fomentar la colaboración entre gobiernos, industria, academia y sociedad civil para asegurar que el desarrollo de la IA sirva como una fuerza para el avance global inclusivo y sostenible.

Las iniciativas educativas que equipan a diversos segmentos de la sociedad con alfabetización y habilidades de gestión de IA son cruciales para nivelar el campo de juego. Por ejemplo, los programas financiados por gobiernos, que se centran en el entrenamiento en IA para trabajadores de ingresos medios y bajos, pueden ser instrumentales en reducir la desigualdad. Tales programas ayudan a los individuos a aprovechar la IA para obtener ganancias en productividad, reduciendo así el riesgo de disparidades económicas exacerbadas.

El impacto de la IA en la desigualdad económica depende significativamente de quién experimenta el desplazamiento laboral versus quién puede aprovechar la IA para impulsar la productividad. Informes, como los del Fondo Monetario Internacional (FMI), sugieren que los trabajadores mejor pagados pueden ver ganancias de ingresos desproporcionadas debido a la complementariedad con la IA, potencialmente aumentando la desigualdad de ingresos laborales. Para abordar esto, las discusiones políticas y foros de grupos de expertos (*think tanks*) podrían explorar estrategias para la distribución equitativa de los beneficios de la IA, enfatizando programas de educación y entrenamiento adaptados a diferentes niveles de habilidad y grupos de edad. Iniciativas como el sistema de educación dual de Alemania, que integra la formación vocacional con la educación formal, podrían servir como modelo para otras naciones.

Los enfoques nacionales para definir los derechos de propiedad de la IA e implementar políticas fiscales redistributivas serán decisivos en dar forma a la influencia de la IA en la distribución de ingresos y riqueza. Si la IA mejora significativamente ciertos trabajos, la mayor demanda de mano de obra calificada podría contrarrestar los efectos de la automatización, potencialmente estabilizando o incluso mejorando la distribución de ingresos. Esto resalta la importancia de talleres de políticas y debates centrados en los derechos de propiedad de la IA y estrategias fiscales dirigidas a formular políticas económicas inclusivas. Orga-

nizaciones globales, como las Naciones Unidas, la iniciativa del Departamento de Ciencia, Innovación y Tecnología (DSIT) del Reino Unido que ha convocado cumbres semestrales en todo el mundo sobre la seguridad de la IA, y la OCDE, podrían liderar en facilitar estas discusiones para asegurar un consenso e implementación generalizada.

La disposición para adoptar y beneficiarse de la IA varía notablemente según el nivel educativo y la edad. Los trabajadores con títulos universitarios y los individuos más jóvenes son típicamente más adaptables y propensos a transitar hacia roles complementarios con la IA. Por el contrario, los trabajadores sin educación post-secundaria y los individuos mayores pueden enfrentar susceptibilidad creciente a los cambios inducidos por la IA. Los programas dirigidos de mentoría y aprendizaje continuo para estos grupos vulnerables pueden ayudar a mitigar los riesgos y promover una transición más inclusiva hacia una economía mejorada por la IA. Por ejemplo, las asociaciones público-privadas y otras versiones menos onerosas del papel del Estado, que patrocinan iniciativas de recapacitación pueden jugar un papel crítico en apoyar estas transiciones.

La disparidad en la adopción y beneficios de la IA entre países subraya la necesidad de invertir en infraestructura digital, capital humano, innovación tecnológica y marcos legales adaptables. Índices como el Índice de Preparación Gubernamental para la IA de la Universidad de Oxford y el Índice de Preparación para la IA del FMI, indican que tanto las economías avanzadas como las emergentes necesitan priorizar la integración de la IA mientras fomentan un entorno regulatorio que permita el crecimiento. Las conferencias internacionales y las iniciativas de investigación colaborativa centradas en compartir mejores prácticas en infraestructura digital y política de IA pueden ayudar a reducir las disparidades y fomentar un avance uniforme a través de las regiones. Ejemplos de colaboraciones exitosas incluyen asociaciones en el Sudeste Asiático que unen recursos para construir capacidades de IA en naciones menos desarrolladas.

Los elementos clave para la adopción efectiva de la IA incluyen inversión sostenida en educación, fomento de experiencias en ciencia, tecnología, ingeniería y matemáticas inclusivas y promoción de la movilidad laboral y de capital. Los marcos legales que acomodan nuevos modelos de negocio digitales también son esenciales. Desarrollar una fuerza laboral digitalmente capacitada, respaldada por infraestructura fundamental e innovación, es crítico para todas las economías. Igualmente importantes son las redes de seguridad social y los programas de recapacitación para trabajadores en riesgo de desplazamiento laboral debido a los avances de la IA, asegurando la inclusividad y previniendo

mayor desigualdad. Proponer asociaciones globales y mecanismos de financiamiento para apoyar estos elementos clave puede ayudar a asegurar que los beneficios de la IA sean ampliamente accesibles.

El retraso en adaptarse al rápido progreso de la IA podría llevar a la inestabilidad e inseguridad social. Implementar servicios básicos universales y considerar medidas como la renta básica universal (RBU) puede ser esencial para reducir la brecha entre diferentes grupos sociales y asegurar la participación equitativa en el futuro impulsado por la IA. Organizar diálogos globales sobre sistemas innovadores de apoyo social, como la RBU, podría proporcionar una plataforma para explorar y pilotear enfoques que mitiguen los impactos socioeconómicos de la IA, apoyando una transición más estable e inclusiva. Por ejemplo, los programas piloto de RBU en países como Finlandia y Canadá han mostrado beneficios potenciales en reducir la inseguridad económica.

La Cumbre de Seguridad de la IA auspiciada por el Reino Unido, en la que participé, sirvió como un duro recordatorio de la urgencia de colaborar en abordar los desafíos arquitectónicos, los numerosos impactos sociales de la IA y la gobernanza regulatoria, antes de que nos enfrentemos a un potencial futuro distópico de nuestra propia creación. Solo a través de la colaboración coordinada, arraigada en principios éticos y una visión compartida de la humanidad, podemos aprovechar el poder de la IA para beneficiar a todos. Los próximos años serán cruciales en dar forma a este camino, exigiendo acción inmediata, colectiva e informada de gobiernos, organismos internacionales, corporaciones y sociedad civil. Por ello es vital la colaboración internacional y la conformación de un sistema regulatorio global y adaptativo que incluya a todos los países.

VI. LA NECESIDAD DE COLABORACIÓN INTERNACIONAL

La gobernanza de la IA necesita un enfoque holístico y global que integre aspectos políticos, económicos, sociales, éticos, de derechos humanos, técnicos y ambientales. Este enfoque integral permitiría transformar el actual mosaico fragmentado de iniciativas, que hasta ahora ha demostrado ser ineficaz y contraproducente, en un sistema coherente e interoperable. Basado en el derecho internacional y alineado con los Objetivos de Desarrollo Sostenible (ODS), esta estrategia debe ser adaptable a distintos contextos y capaz de evolucionar con el tiempo.

El rápido desarrollo de la IA ha llevado al surgimiento de numerosos estándares adoptados por organizaciones internacionales como la Unión Internacional

de Telecomunicaciones (UIT), la Organización Internacional de Normalización (ISO), la Comisión Electrotécnica Internacional (IEC) y el Instituto de Ingenieros Eléctricos y Electrónicos (IEEE). Si bien esta proliferación internacional, y sus ecos nacionales y hasta locales, ha ayudado a crear pautas fundamentales, también ha resultado en un panorama fragmentado sin un lenguaje común o marco unificado. La ausencia de definiciones acordadas para conceptos clave como equidad, seguridad y transparencia complica aún más los esfuerzos de gobernanza cohesiva, llevando a discrepancias entre los estándares técnicos estrechamente definidos y aquellos destinados a incorporar principios éticos más amplios.

Uno de los beneficios más inmediatos de la cooperación internacional en la gobernanza de la IA es la combinación de experiencia y recursos. Los países con capacidades avanzadas en IA pueden compartir conocimientos, estándares técnicos y mejores prácticas con naciones que pueden carecer de la misma infraestructura tecnológica o experiencia. Este intercambio no solo acelera la innovación, sino que también ayuda a crear una base para estándares éticos y técnicos que benefician a todas las partes involucradas. Por ejemplo, las iniciativas conjuntas de investigación y los programas educativos transfronterizos pueden ayudar a las naciones en desarrollo a construir competencias en IA, fomentando un panorama tecnológico global más equilibrado.

Los estándares internacionales consistentes pueden simplificar el despliegue de tecnologías de IA y reducir la fricción regulatoria para las empresas multinacionales. Las regulaciones armonizadas aseguran que las aplicaciones de IA se adhieran a pautas éticas compartidas y protocolos de seguridad, minimizando el riesgo de prácticas dañinas o no éticas. Esta alineación también apoya a las empresas proporcionando un marco más claro dentro del cual pueden operar, reduciendo los costos de cumplimiento y fomentando el crecimiento económico.

La naturaleza global de los riesgos potenciales de la IA, como las armas autónomas o los ciberataques avanzados, necesita acción colectiva inmediata y efectiva. La cooperación internacional permite el desarrollo de protocolos conjuntos para manejar crisis relacionadas con la IA y refuerza la seguridad global al prevenir el mal uso de las tecnologías de IA. Los datos compartidos sobre vulnerabilidades de la IA y los esfuerzos coordinados para abordarlas pueden ayudar a los países a adelantarse a amenazas emergentes, y a interceptar y neutralizar a actores maliciosos.

Los marcos de gobernanza colaborativa promueven la transparencia y la rendición de cuentas, que son críticas para construir la confianza pública. Las

políticas conjuntas que reflejan valores humanos compartidos alientan a los ciudadanos a ver los avances de la IA como alineados con sus intereses. La percepción pública de la legitimidad y seguridad de la IA mejora cuando las políticas están informadas por una gama de perspectivas globales e incluyen controles contra sesgos y consecuencias no intencionadas.

Un desafío importante en la gobernanza actual de la IA es la falta de coordinación entre varias iniciativas e instituciones, incluyendo aquellas dentro de las Naciones Unidas. Este enfoque desarticulado arriesga crear estructuras de gobernanza fragmentadas e inefectivas, donde los marcos dispares pueden entrar en conflicto o fallar en abordar los problemas críticos de manera integral. Sin esfuerzos cohesivos para alinear las estrategias globales, el mundo podría enfrentar regímenes de gobernanza de IA desconectados e incompatibles que socavan la interoperabilidad, obstaculizan las respuestas transfronterizas a incidentes relacionados con la IA y, en última instancia, debilitan los beneficios compartidos del progreso tecnológico.

VII. DESAFÍOS DE LA COLABORACIÓN INTERNACIONAL

Lograr la convergencia entre las principales potencias globales –particularmente Estados Unidos, China y la Unión Europea– presenta desafíos sustanciales debido a las estructuras políticas y filosofías regulatorias divergentes. Estados Unidos a menudo adopta un enfoque impulsado por el mercado que prioriza la innovación y el crecimiento económico, mientras que la UE enfatiza protecciones estrictas de privacidad y pautas éticas. China, por otro lado, integra el control estatal y la supervisión generalizada en su enfoque, centrándose en la seguridad nacional y la influencia económica. Estas diferencias pueden crear fricciones y obstaculizar los esfuerzos para acordar principios compartidos para la gobernanza de la IA.

El valor estratégico de la IA como herramienta para lograr una ventaja económica y militar amplifica la competencia entre las principales potencias. Esta rivalidad puede llevar a políticas protectoras o exclusivas que priorizan los intereses nacionales sobre la colaboración global. Por ejemplo, las disputas comerciales en curso, visiblemente entre Estados Unidos y China, y las rivalidades tecnológicas de diversos órdenes entre países avanzados, pueden impedir el diálogo abierto y el intercambio de datos. Sin mecanismos para aliviar estas tensiones, los marcos cooperativos corren el riesgo de verse socavados por la desconfianza mutua y las agendas geopolíticas.

La cooperación internacional también debe abordar las disparidades entre las naciones tecnológicamente avanzadas y aquellas que aún están desarrollando sus capacidades en IA. Sin apoyo específico, las naciones menos desarrolladas pueden encontrarse incapaces de participar significativamente en las discusiones de gobernanza, ampliando la brecha tecnológica global. Este desequilibrio podría resultar en regulaciones que no tomen en cuenta diversos contextos económicos y culturales, marginando aún más a ciertas regiones.

Incluso cuando se alcanza el consenso, implementar y hacer cumplir las regulaciones internacionales puede ser difícil. El sistema legal y político de cada país afecta cómo se integran los acuerdos en las leyes nacionales. Además, monitorear y asegurar el cumplimiento de los estándares internacionales requiere recursos y supervisión coordinada, lo que puede ser difícil de mantener a lo largo del tiempo.

La gobernanza global de la IA debe ir más allá de los acuerdos y principios y enfocar una implementación que sea viable. Asegurar que los compromisos globales se traduzcan en resultados tangibles es crucial para un progreso significativo. Esto incluye iniciativas para el desarrollo de capacidades, particularmente en regiones con acceso limitado a recursos de IA, y apoyo para pequeñas y medianas empresas (PYME) para que puedan competir e innovar efectivamente. Sin medidas efectivas que traduzcan la política en práctica, la gobernanza global corre el riesgo de convertirse en un ejercicio puramente ceremonial, desprovisto de impacto real.

VIII. EL CAMINO A SEGUIR

La colaboración internacional es crucial para desarrollar un marco global que regule la IA alineándola con valores humanos y estándares de seguridad. Aunque organizaciones como la ONU y la OCDE pueden facilitar el diálogo y establecer principios comunes, la convergencia regulatoria entre potencias principales enfrenta desafíos debido a sus diferentes estructuras políticas e intereses estratégicos. No obstante, esta cooperación es esencial para asegurar una distribución equitativa de los beneficios de la IA y una gestión colectiva de sus riesgos, permitiendo que la tecnología sirva al bien común de la humanidad.

Comprender estas fuerzas impulsoras y desafíos es esencial para evaluar las oportunidades y riesgos asociados con la IA. Ignorar estas dinámicas plantea amenazas significativas para el desarrollo responsable de la IA, incluyendo problemas relacionados con la participación y la distribución equitativa de beneficios. La investigación científica continua y los esfuerzos interdisciplinarios

son vitales para asegurar que las tecnologías de IA permanezcan compatibles con los derechos humanos, los valores sociales y el objetivo más amplio de avanzar en el bienestar humano. Cerrar las brechas tecnológicas y de políticas a través de la gobernanza proactiva y la cooperación internacional determinará en última instancia si la IA sirve como una poderosa herramienta para el progreso humano o una fuente de profundos desafíos sociales.

La guía para la formación de nuevas instituciones internacionales de gobernanza de IA debe basarse en principios que reconozcan el contexto más amplio en el que opera la IA. Estos incluyen reconocer que la gobernanza de la IA no ocurre en el vacío y debe estar alineada con el derecho internacional existente, particularmente el derecho internacional de los derechos humanos. Esta alineación asegura que el desarrollo y despliegue de la IA promuevan la dignidad humana, la equidad y la protección de los derechos fundamentales.

GOBERNANZA REGULATORIA ADAPTATIVA DE LA IA: HACIA UN FUTURO COMPATIBLE, CONTENIDO Y ALINEADO

El ritmo exponencial del desarrollo de la IA ha superado la capacidad de los marcos regulatorios tradicionales, creando un desafío significativo para la gobernanza efectiva. Como se mencionó antes, el panorama actual se caracteriza por numerosos borradores de leyes relacionadas con la IA, críticas generalizadas a las respuestas gubernamentales lentas y percepciones polarizadas que enmarcan la IA como una puerta hacia la utopía o un camino hacia la distopía. Este entorno complejo ha llevado a la confusión, el estancamiento y un riesgo elevado. Superar estos desafíos requiere un enfoque adaptativo y resiliente para la gobernanza –uno que evolucione junto con la IA, alineando el avance tecnológico con las prioridades y valores sociales.

La IA tiene el potencial de fortalecer los derechos humanos, el bienestar y las funciones sociales. Este potencial subraya la importancia de un enfoque de gobernanza que sea proactivo, basado en evidencia y receptivo. La ciencia y la experiencia interdisciplinaria deben impulsar esta evolución regulatoria, apuntando a crear un marco de gobernanza que equilibre la gestión de riesgos, la regulación, la inversión y la innovación. Tal marco debe fomentar medidas inteligentes y adaptativas, permitiendo incentivos para la inversión empresarial y la innovación mientras mantiene una base ética rigurosa. Los esfuerzos existentes, como la Ley de IA de la Unión Europea, demuestran intentos de establecer clasificaciones y marcos de IA de alto riesgo, y es posible que se constituyan en un punto de partida para la adaptación global futura.

La gobernanza efectiva de la IA requiere equilibrar la mitigación de riesgos con el fomento de la innovación. Este equilibrio necesita un enfoque regulatorio que fomente la inversión y el emprendimiento mientras se adhiere a principios globales, como los establecidos por la Carta de las Naciones Unidas y la Declaración Universal de los Derechos Humanos. Sin embargo, esta tarea se complica por las variadas estrategias regulatorias de los principales actores globales mencionadas antes. Las prioridades políticas, económicas y culturales únicas de cada región contribuyen a un entorno competitivo que desafía la colaboración internacional cohesiva en la gobernanza de la IA.

El estado actual de la gobernanza de la IA tiene paralelos con aspectos del concepto de la Guerra Fría de "destrucción mutua asegurada" (MAD, por sus siglas en inglés), pero con características e implicaciones distintas. La IA tiene el potencial de devastar a la civilización humana, como el uso de las armas nucleares. Sin embargo, a diferencia de la tecnología nuclear, la IA es una fuerza cognitiva en evolución, capaz de automejora y del desarrollo de sistemas complejos y opacos, que potencialmente podrían escapar al control humano. Esto introduce un nuevo concepto, "camino hacia la destrucción asegurada" (PAD), donde el desarrollo no controlado de la IA plantea riesgos existenciales inminentes. Estos riesgos incluyen la creación de armamento autónomo con consecuencias impredecibles, haciendo esencial la cooperación global. Los esfuerzos recientes de organismos como UNESCO y la OCDE, que han propuesto pautas éticas para la IA, proporcionan marcos valiosos que podrían contribuir a forjar principios universales. Sin embargo, estos necesitan traducirse en estándares ejecutables a través de acuerdos globales.

Progresar hacia la compatibilidad, contención y alineación de la IA necesita un compromiso con principios universales que guíen el desarrollo de la gobernanza adaptativa. Asegurar que ningún desarrollo de superinteligencia artificial (ASI) avance sin evidencia incontrovertible de su controlabilidad y beneficios para la supervivencia humana es un componente clave. Esto implica evaluaciones integrales de riesgos y la implementación de salvaguardias que prioricen el bienestar de la humanidad. También es esencial integrar el compromiso público y la transparencia en los esfuerzos de gobernanza, fomentando la confianza y asegurando que diversas perspectivas sociales configuren la política.

Fomentar la innovación en IA debe ocurrir dentro de marcos de supervisión humana y responsabilidad legal, asegurando que las decisiones y acciones de la IA siempre permanezcan atribuibles a operadores humanos. Este enfoque apoya la transparencia y salvaguarda contra prácticas no éticas, negligentes o dañinas en el desarrollo y uso de la IA, haciendo cumplir pautas estrictas y

consecuencias legales para los infractores. Ejemplos de esto incluyen entornos regulatorios de prueba que permiten la confirmación y el desarrollo controlado de aplicaciones de IA, facilitando fomentar y conducir la innovación mientras se monitorean los riesgos potenciales. Combatir el robo de información digital y las prácticas no éticas también requiere medidas robustas de ciberseguridad, marcos legales claros y cooperación internacional para proteger los derechos digitales individuales y colectivos.

Para fomentar la confianza y la responsabilidad, el papel de la IA en la creación de contenido debe estar regido por pautas claras que aseguren que el contenido generado por IA sea identificable, rastreable y consistente con los principios de libre expresión. Esto incluye prevenir que la IA suplante a individuos u organizaciones sin consentimiento explícito e informado, salvaguardando así la identidad y la autenticidad. Se necesita un diálogo global para negociar con las corporaciones de IA, instándolas a asignar una parte de sus ganancias hacia la aplicación de estándares, la mitigación de impactos socioeconómicos y el apoyo a las comunidades afectadas. Iniciativas como asociaciones público-privadas y diálogos con las partes interesadas pueden facilitar tales acuerdos y promover el progreso cooperativo.

Promover el uso responsable de la IA en abordar la seguridad pública, los derechos humanos y los derechos de propiedad intelectual es crucial, al igual que asegurar la alineación con los valores sociales y las normas legales. El establecimiento de mecanismos para monitorear, intervenir y procesar a aquellos que usan indebidamente la IA con propósitos hostiles o negligentes refuerza los estándares legales y éticos. Igualmente, clarificar y hacer cumplir la responsabilidad legal y la rendición de cuentas concerniente a las acciones de la IA asegura que existan pautas claras para la reparación y compensación cuando sea necesario. El compromiso público, a través de consultas y toma de decisiones participativa, puede ayudar a dar forma a estas políticas para reflejar los valores y preocupaciones sociales.

La adopción universal de estos principios apunta a fomentar un enfoque equilibrado para el desarrollo de la IA, minimizando los riesgos existenciales mientras se prioriza la seguridad humana y se fomentan las prácticas éticas. Este enfoque busca alentar la innovación mientras se abordan los desafíos potenciales asociados con las tecnologías avanzadas de IA. Al construir un sistema de gobernanza flexible y adaptativo arraigado en la cooperación internacional y la supervisión ética, las sociedades pueden aprovechar el potencial de la IA para un beneficio generalizado mientras se protegen contra sus riesgos. El camino por delante requiere un compromiso inquebrantable de los gobiernos, el sector

privado y la sociedad civil para crear un futuro donde la IA avance el bienestar humano sin comprometer la seguridad o los estándares éticos.

HACIA UN SISTEMA DE GOBERNANZA REGULATORIA ADAPTABLE

Mientras la adaptabilidad institucional y regulatoria progresa incrementalmente, el cambio tecnológico –especialmente en Inteligencia Artificial de Propósito General– avanza a un ritmo exponencial. Esta creciente disparidad subraya la necesidad urgente de enfoques innovadores de gobernanza que puedan evolucionar en paralelo con el rápido desarrollo de la IA. Sin sistemas regulatorios adaptables y receptivos, la brecha entre el ritmo de la evolución tecnológica y la supervisión regulatoria continuará ampliándose, reduciendo la efectividad de la gestión estratégica y organizacional. Este desajuste presenta un desafío crítico que debe abordarse para prevenir la obsolescencia regulatoria y salvaguardar los intereses sociales.

La necesidad apremiante de un sistema regulatorio adaptable que pueda responder en tiempo real a la incesante emergencia de nuevas tecnologías de IA se está volviendo cada vez más evidente. Si las estructuras de gobernanza permanecen estáticas, o sólo avanzan gradualmente, corren el riesgo de volverse obsoletas a un ritmo acelerado, lo que podría llevar a consecuencias imprevistas con implicaciones globales potencialmente significativas. Es poco probable que los marcos regulatorios tradicionales impulsados por humanos, actualmente en vigor, mantengan el ritmo con el crecimiento exponencial de la IA y las tecnologías relacionadas. Este retraso significa que las reformas bien intencionadas y las pautas éticas pueden volverse inefectivas, o incluso contraproducentes, si no pueden adaptarse lo suficientemente rápido al panorama cambiante.

Las sinergias tecnológicas podrían jugar un papel fundamental en superar estos desafíos. Las tecnologías emergentes como *blockchain*, aprendizaje automático y análisis de datos avanzados ofrecen herramientas poderosas para mejorar la gobernanza adaptativa. Por ejemplo, *blockchain* puede proporcionar transparencia y trazabilidad en los procesos regulatorios, mientras que los sistemas de monitoreo impulsados por IA pueden señalar el incumplimiento en tiempo real, permitiendo acciones correctivas más rápidas y asegurando la adherencia a las regulaciones. Estas herramientas podrían contribuir a construir un sistema de gobernanza receptivo capaz de supervisión continua y adaptación rápida a nuevos desarrollos.

Sin embargo, implementar un sistema de gobernanza adaptable no está exento de desafíos significativos. La resistencia de instituciones establecidas, que pueden ser reacias a revisar las prácticas regulatorias tradicionales, es un obstáculo común. Además, los costos asociados con la actualización y mantenimiento continuo de un marco adaptativo pueden ser considerables, planteando preocupaciones presupuestarias y logísticas para muchos gobiernos y organizaciones. Otro problema apremiante es el riesgo de captura regulatoria, donde las empresas tecnológicas poderosas ejercen una influencia indebida sobre el proceso regulatorio para dar forma a resultados que favorezcan a sus intereses. Abordar estos desafíos requiere un enfoque equilibrado que asegure la integridad e imparcialidad de las estructuras de gobernanza, mientras se fomenta la colaboración entre los sectores público y privado.

Para gestionar y contener el desarrollo y despliegue de la IA de una manera segura, ética y responsable, existe una necesidad urgente de un sistema de gobernanza regulatoria que sea dinámico, flexible y altamente adaptable. Tal sistema debe ser capaz de evolucionar al paso de los avances tecnológicos. El desarrollo de este sistema de gobernanza debe involucrar participación global, aprovechando las contribuciones de gobiernos, empresas privadas, instituciones académicas, sociedad civil y el público general. Este enfoque inclusivo asegura que el sistema sea transparente, responsable y caracterizado por mecanismos claros para el monitoreo, retroalimentación y modificación. La versatilidad es clave para responder efectivamente a los cambios rápidos y a menudo impredecibles en la tecnología.

En alineación con estos principios, cualquier organismo regulador o sistema diseñado para gestionar la supervisión estratégica de la IA debe cumplir roles esenciales. Primero, debe monitorear y guiar continuamente el desarrollo e implementación de tecnologías emergentes de IA para mantener una comprensión continua de sus impactos y riesgos potenciales. Segundo, el proceso regulatorio debe ser eficiente, confiable e inclusivo, asegurando que todas las partes interesadas relevantes estén involucradas y representadas. Tercero, este sistema debe estar facultado para bloquear la progresión de tecnologías de IA que presenten riesgos significativos o puedan llevar a resultados negativos, actuando así como una salvaguarda para el bienestar público. Finalmente, fomentar el compromiso y la cooperación global es crucial para compartir conocimiento, alinear estándares y aprovechar la experiencia colectiva para mitigar riesgos y maximizar beneficios compartidos.

Para que la estructura de gobernanza regulatoria supervise efectivamente el desarrollo de la IA, debe exhibir la flexibilidad para adaptarse a cambios

rápidos e impredecibles. Una estrategia exitosa aseguraría que la supervisión regulatoria no sofoque la innovación, sino que, en cambio, fomente un entorno donde el crecimiento tecnológico se alinee con principios éticos y sirva al interés público global. Lograr este equilibrio requiere un enfoque colaborativo, donde diversas voces contribuyan al proceso de toma de decisiones, asegurando que los resultados sean equitativos y que los beneficios del progreso tecnológico se compartan entre países y a lo largo de la sociedad.

Desarrollar tal sistema de gobernanza requerirá la integración de herramientas digitales y la IA misma para apoyar e impulsar prácticas regulatorias adaptativas. El sistema debe ser capaz de realizar análisis de datos en tiempo real, planificación de escenarios y ajuste dinámico de políticas para mantener su relevancia. Además, la cooperación internacional debe ser priorizada para crear regulaciones estandarizadas que sean lo suficientemente flexibles para acomodar diferentes contextos regionales y culturales mientras mantienen principios fundamentales que protegen los derechos humanos y promueven el uso ético de la IA.

Para abordar las inconsistencias y cerrar la brecha entre diferentes estándares de IA, se podría establecer una cámara de compensación central para la gobernanza de la IA bajo el sistema de las Naciones Unidas. Este organismo reuniría experiencia de paneles científicos internacionales, organizaciones de estándares nacionales e internacionales, empresas tecnológicas y representantes de la sociedad civil. Tal plataforma podría ayudar a alinear los estándares técnicos y éticos globalmente, asegurando que la gobernanza de la IA sea coherente, integral y fundamentada en valores compartidos.

Un enfoque unificado para la estandarización de la IA no solo simplificaría la gobernanza, sino que también apoyaría el desarrollo equitativo y transparente. La participación de diversas partes interesadas en este proceso es crucial para promover la confianza y la responsabilidad. Al actuar como una cámara de compensación, la ONU podría facilitar la alineación de estándares existentes y crear un entendimiento común que apoye tanto la excelencia técnica como la responsabilidad ética.

El siguiente paso involucra solidificar estos marcos a través de programas piloto, acuerdos transfronterizos y legislación adaptativa que pueda ajustarse a medida que emergen nuevos datos y resultados. Esto requerirá compromiso de todas las partes interesadas y reconocimiento de la complejidad involucrada en adaptar la gobernanza. Solo adoptando tales medidas puede la comunidad global navegar los desafíos planteados por la IA, asegurando que su desarrollo continúe beneficiando a la humanidad sin socavar los valores sociales o la seguridad.

Un obstáculo mayúsculo es ir más allá de la competencia geopolítica y la desconfianza mutua entre las potencias tecnológicas, que agravan sus diferencias. Es vital acordar medidas de construcción de confianza y cooperación internacional. Pese a las dificultades, hay oportunidades para un progreso incremental a través de la colaboración en áreas de interés común, como la seguridad y la ética de la IA. El papel de plataformas internacionales, como las Naciones Unidas, la iniciativa de la Cumbre de Seguridad de la IA, y la OCDE, son cruciales para facilitar el diálogo y establecer estándares compartidos.

Para avanzar, la gobernanza de la IA debe ser flexible, adaptativa y centrada en el ser humano, con principios de transparencia y rendición de cuentas que refuercen la confianza pública. La inclusión de países en desarrollo, del llamado Sur Global, y la implementación de mecanismos de monitoreo efectivos garantizarán que los beneficios de la IA se distribuyan de manera equitativa, evitando la exacerbación de desigualdades globales y, con ello, los terribles impactos sobre la inseguridad e incertidumbre globales. Un enfoque colaborativo, que integre a gobiernos, sector privado, academia y sociedad civil, es esencial para lograr una gobernanza que equilibre innovación y seguridad en el contexto de la evolución tecnológica acelerada. Todos deben poder contribuir, participar y beneficiarse de estos esfuerzos.

Una conclusión inevitable es que, si la humanidad no se une para evitar los riesgos catastróficos de una IA descontrolada y en manos de actores malintencionados, enfrentará irremediablemente peligros existenciales que podrían condenar la nueva era en la historia de la humanidad a una catástrofe existencial.

LA NECESIDAD DE UN SISTEMA DE GOBERNANZA REGULATORIA ADAPTATIVO RESPALDADO POR IA

Ante los crecientes desafíos y riesgos del avance de la inteligencia artificial (IA), se hace indispensable un sistema de gobernanza adaptativo respaldado por IA que pueda gestionar eficazmente su desarrollo y evolución. La integración de la IA en la gobernanza va más allá de la supervisión tradicional, ofreciendo herramientas que permiten una respuesta dinámica a un entorno tecnológico en constante cambio. Adoptar marcos adaptativos asegura que las medidas regulatorias puedan mantenerse al ritmo de la IA, mitigando los riesgos y maximizando los beneficios.

La IA tiene la capacidad de transformar la adaptabilidad de los marcos regulatorios al ofrecer respuestas ágiles y basadas en datos. Al analizar grandes

volúmenes de información, como indicadores económicos y tendencias sociales, los sistemas de IA pueden identificar patrones emergentes y cambios en el sentimiento público o en las condiciones del mercado. Esta capacidad predictiva es invaluable para actualizar políticas y regulaciones, asegurando su relevancia y eficacia. Por ejemplo, herramientas de cumplimiento impulsadas por IA se utilizan en mercados financieros para detectar irregularidades y gestionar riesgos en tiempo real, mejorando la capacidad de respuesta de los reguladores.

En el ámbito del análisis legal, la IA puede evaluar marcos normativos existentes, identificar inconsistencias y sugerir actualizaciones para armonizar regulaciones entre distintas regiones. Al simular el impacto de nuevas regulaciones bajo escenarios hipotéticos, la IA ayuda a revelar vacíos en las estructuras legales y facilita un enfoque proactivo en la adaptación legislativa. Además, los sistemas de IA pueden asistir en la redacción de documentos legales alineados con leyes y precedentes actuales, garantizando coherencia y precisión.

La capacidad de monitoreo de la IA es igualmente esencial en la supervisión del cumplimiento a través de industrias. Al detectar patrones de incumplimiento y prever áreas problemáticas, la IA permite a los organismos reguladores actuar preventivamente. Esto optimiza la asignación de recursos, permitiendo que los reguladores se concentren en la toma de decisiones estratégicas. En la regulación financiera, por ejemplo, la IA se ha utilizado para identificar actividades sospechosas y prevenir fraudes, fortaleciendo la seguridad y efectividad del sistema regulatorio.

La IA también puede proporcionar información sobre el sentimiento público mediante el análisis de redes sociales y discursos en línea. Al identificar tendencias y preocupaciones, los formuladores de políticas pueden ajustar las regulaciones para alinearlas con las expectativas sociales y mejorar la representatividad de las políticas. Este ciclo de retroalimentación en tiempo real refuerza la transparencia y la confianza pública, asegurando que las políticas evolucionen en sintonía con las necesidades de los ciudadanos.

Sin embargo, para garantizar que la IA opere de manera justa y sin sesgos, es fundamental implementar auditorías y actualizaciones regulares. Esto incluye el uso de datos diversos y la aplicación de algoritmos de detección de sesgos para asegurar estándares equitativos. Deben establecerse mecanismos claros para la interpretación de decisiones generadas por IA, manteniendo la responsabilidad final en manos de los tomadores de decisiones humanos. La IA debe ser una herramienta de apoyo, no un sustituto, garantizando así la rendición de cuentas y el alineamiento con valores democráticos y derechos humanos.

La adopción de la IA en la gobernanza enfrenta retos, como la resistencia de instituciones y formuladores de políticas que pueden temer la falta de transparencia o el desplazamiento laboral. Para superar estas barreras, es necesario comunicar claramente los beneficios de la IA, implementar programas piloto y ofrecer capacitación para demostrar su valor en la gestión regulatoria.

El compromiso con principios éticos es innegociable en la integración de la IA. La transparencia, la equidad y la rendición de cuentas deben guiar la evolución de estos sistemas. Las evaluaciones periódicas y la adaptación a los avances tecnológicos son cruciales para evitar la obsolescencia de la IA en la gobernanza. Un enfoque de gobernanza adaptativa respaldado por IA también puede influir en la cooperación internacional, sirviendo de modelo para acuerdos regulatorios y promoviendo la estandarización global de políticas.

Organizaciones internacionales como el Organismo Internacional de Energía Atómica y el Reactor Termonuclear Experimental Internacional han demostrado que es posible gestionar tecnologías complejas con mandatos autorizados y una toma de decisiones inclusiva. Estos ejemplos pueden inspirar enfoques similares en la gobernanza de la IA, siempre y cuando se incluya al Sur Global en los procesos para asegurar una representación equitativa.

La capacidad de la IA para respaldar un sistema regulatorio adaptativo que evolucione al ritmo de la tecnología puede empoderar a las estructuras de gobernanza, fomentando la confianza mutua y sirviendo como punto de referencia para estándares globales unificados.

El costo de no implementar un sistema de gobernanza regulatoria adaptativo es alto: una carrera tecnológica descontrolada, la ampliación de la brecha digital y la erosión de la confianza pública. Es esencial establecer principios universales que prioricen los derechos humanos, la seguridad y el bienestar, junto con marcos regulatorios flexibles que guíen el desarrollo de la IA hacia el bien común.

LIBERALISMO, DEMOCRACIA E INTELIGENCIA ARTIFICIAL. LOS RETOS DE LA LIBERTAD

SUMARIO

I. Introducción. II. ¿Democracia enferma? III. ¿El liberalismo agotado? IV. Liberalismo, democracia y algoritmos. **V.** Hacia una utopía de lo digital. **VI.** De datologías y distopías. **VII.** Regular o no regular. That's the question? **VIII.** Reflexiones generales. IX. Bibliografía.

> ARTURO OROPEZA GARCÍA

Doctor en Derecho e Investigador Nacional del Instituto de Investigaciones Jurídicas de la Universidad Nacional Autónoma de México (IIJ, UNAM). Presidente del Instituto para el Desarrollo Industrial y la Transformación Digital A.C. (INADI).

I. INTRODUCCIÓN

Desde finales del siglo XX y a lo largo del presente siglo, un clamor que ha ido subiendo en intensidad y en desconcierto ha estado ocupando la preocupación de los especialistas, de los políticos y de la sociedad en general.

La sensación de que en la política occidental algo no va bien ha rebasado en muchos países la línea de lo tolerable. No son pocas las voces ni los movimientos que en medio de esta preocupación han iniciado la búsqueda de las causas de una *democracia enferma*.

Desde diferentes plataformas se habla de *El ocaso de la democracia* (Applebaum); *Cómo mueren las democracias* (Levitsky y Ziblatt); *La caída y ascenso de la democracia* (Stasavage); *Una teoría de la democracia compleja (*Innerarity); *Los linderos del caos* (Moyo); *El pueblo contra la democracia* (Mounk), y muchos otros que nos hablan de una enfermedad creciente (déficit, recesión, agotamiento, etc.) de la sociedad occidental de nuestro tiempo.

Otro grupo de especialistas no se limitan solamente a llamar la atención de un déficit democrático. Ampliando la mirada nos hablan de *El liberalismo y sus desencantados* (Fukuyama); De *la crisis del capitalismo democrático* (Wolf); Del *postcapitalismo* (Mason); *Del ascenso y la caída del orden neoliberal* (Gerstle); *De un mundo después del liberalismo* (Rose); De *el liberalismo herido* (Lassalle), etc., subrayando que el problema político del siglo XXI no se limita a un déficit de democracia, sino que la enfermedad política de la sociedad occidental escala hacia un modelo explicativo filosófico, político, económico y social que la orientó y le dio rumbo a lo largo de casi tres siglos.

Por otro lado, en una pequeña bibliografía que empieza a surgir, autores como Suleyman (*La ola que viene*), Lassalle (*Ciberleviatán*), Zuboff (*La era del*

capitalismo de la vigilancia), Acemoğlu y Johnson (*Power and progress*), Susskind (*The Digital Republic*), Noah (*21 lecciones para el siglo XXI*), Sadin (*La inteligencia artificial o el desafío del siglo*), Crawford (*Atlas of AI*) y otros, inician la construcción de un análisis más acuciante sobre las nuevas dificultades que se están sumando a la enfermedad democrática y a la crisis del liberalismo, ante la posible aparición de una *infocracia* (Byung-Chul Han)[1], un *ciberleviatán* (Lassalle), un *capitalismo de la vigilancia* (Zuboff), etc., que como una medusa de mil cabezas, a través de su enredadera digital va cubriendo poco a poco la vida política, económica y social del ser humano.

Los retos no son fáciles. Junto a enfermedades endémicas de siempre y las recientes de finales del siglo XX y principios del XXI, propias de un proceso sociopolítico en permanente corrección, surge disruptivamente una poderosa fuerza digital que cuestiona los litigios políticos de hoy, al propio tiempo que convoca a la sociedad global a reflexionar sobre la aparición de una nueva época tecnológica que entre sus múltiples manifestaciones presenta la posibilidad de desaparecer políticamente al ejercicio democrático y a la sociedad que lo contiene, en un futuro de mediano plazo que está más cerca de lo que se le supone.

A continuación se presentan algunas reflexiones generales sobre el tema.

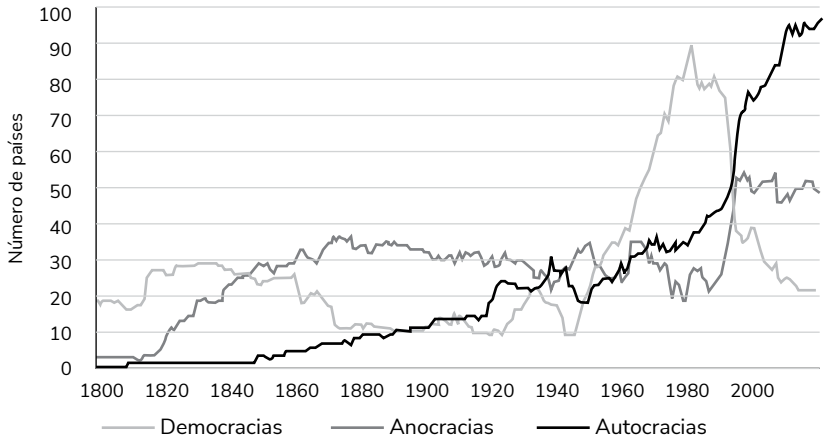
II. ¿DEMOCRACIA ENFERMA?

En el horizonte de la historia de la humanidad, a pesar de la importancia que se le reconoce, la democracia aparece como un fenómeno sociopolítico de nuevo cuño, donde todavía en 1800 no había ningún país que cumpliera con los imperativos democráticos. Cien años después, en 1900, solo había 12 naciones que clasificaban como democráticas y solo una de ellas, Nueva Zelandia, era la que contaba con sufragio universal. Es a lo largo del siglo XX, o sea apenas hace 100 años, cuando la democracia empezó a participar de manera más activa en

1 Llamamos "régimen de la información" a la forma de dominio en la que la información y su procesamiento mediante algoritmos e inteligencia artificial determinan de modo decisivo los procesos sociales, económicos y políticos. A diferencia del régimen de la disciplina, no se explotan cuerpos y energías, sino información y datos. El factor decisivo para obtener el poder no es ahora la posesión de medios de producción, sino el acceso a la información, que se utiliza para la vigilancia psicopolítica y el control y pronóstico del comportamiento. El régimen de la información está acoplado al capitalismo de la información, que hoy deviene en un capitalismo de la vigilancia y que degrada a las personas a la condición de datos y ganado consumidor. (Han, Byung-Chul. (2022), *Infocracia*, Penguin Random House, México, p.9.)

la vida política y social de Occidente. En 1922 se eleva a 24 naciones y en 1989 duplica su número a 48 países, llegando a 2016 a ser parte de la realidad política de aproximadamente la mitad de los países del mundo (97)[2].

GRÁFICA 1 Tendencias globales en modelos de gobierno



FUENTE: Center for Systemic Peace, datos de Polity IV.

A pesar de este éxito aparente, se argumenta que desde el inicio del siglo no menos de 25 democracias se han declarado fallidas alrededor del mundo, incluyendo a Rusia, Turquía y Hungría, como un ejemplo de lo que se padece desde la misma Europa[3].

Freedom House por su lado, reporta que se han registrado 16 años consecutivos de declinamiento de las libertades; y que a la fecha el 80% de la población mundial vive en países autoritarios o parcialmente libres[4]

Un estudio publicado en 2020 concluyó que la ciudadanía democrática ascendía a mil novecientos millones de personas, de las cuales solo algo más del 2% vivía en naciones donde el 75% de los ciudadanos estaban satisfechos con su democracia. El 21% vivía en naciones donde la mitad y las tres cuartas partes estaban satisfechos; y el 57%, o sea la mayoría, vivía en países donde sólo entre un 25 y 50% mostraba satisfacción por vivir en democracia. El resto del 20% habitaba con un 25% o menos de satisfacción democrática[5].

2 Wolf, Martín (2024), *La crisis del capitalismo democrático*, Arial, p.66.
3 Luce, Edward (2017), *The retreat of Western Liberalism*, Atlantic Monthly Press, p.12.
4 Stiglitz, Joseph (2024), The road to freedom, Norton Company, p.1.
5 Wolf, Martín (2024), Op. Cit. p. 71.

En las últimas tres décadas, comentan Foa y Mounk, en las democracias de Occidente ha disminuido drásticamente la confianza en instituciones políticas como los parlamentos o los tribunales. También se ha reducido la participación electoral y la confianza en los partidos políticos; siendo los electores cada vez más proclives a movimientos extremos que votan a candidatos populistas o apoyan a partidos antisistema, mostrando un fuerte deterioro de la democracia occidental[6].

Este sentimiento ha estado acompañado de una idea de derrota, de un alejamiento de la propuesta de un nivel de vida creciente, compartido y sostenible; de la creación de buenos empleos; de igualdad de oportunidades; de mejor seguridad; de una mayor libertad económica y política que redujera las desigualdades y que tuviera un mayor control sobre las grandes concentraciones económicas a través de un Estado que proveyera lo necesario.

De que la democracia está enferma no hay duda de ello. Donde todavía no hay coincidencia es en la naturaleza principal de su padecimiento, para de ahí intentar la reconstrucción de los andamiajes sociales que se hubieran oxidado con el paso del tiempo.

Rose, por ejemplo, no duda en señalar que vivimos un momento pos-neoliberal y sus razones de fondo las ubica en la pérdida de la libertad individual, en la excesiva participación del gobierno y en el cuestionamiento del mercado. También habla de un gran desacuerdo en temas de raza, religión, economía y estrategia política; agregando los aspectos de inmigración, cambio democrático, crecimiento económico, el colapso de la autoridad religiosa, feminismo, igualdad matrimonial, justicia racial, multiculturalismo, etcétera[7].

Bajo una visión más economicista donde se parte del supuesto de que el crecimiento económico es el mejor adhesivo de la democracia liberal, Luce fundamenta la crisis política en el detente económico sufrido por las clases medias, a quienes señala como las principales afectadas del mal comportamiento económico desde la década de los 90's, ocasionando lo que en Francia se ha denominado como la aparición de los *coaches moyennes* (las capas medias afectadas); en Inglaterra como los *left behind* (abandonados), y en América los *squeezed middle* (clase media oprimida); agrupando a todos ellos bajo el término de *precariat* (inestable/limitado). Reconoce que la modernidad ha reducido la pobreza de la humanidad, pero reitera que la clase media, entre una mitad y dos terceras partes, ha sido sacrificada[8].

6 *Ibidem*, p. 70.

7 Rose, Matthew (2021), *A world after liberalism*, Yale, pp.2-4.

8 Luce, Edward (2017), Op. Cit., pp.12-13.

Mounk también arriesga una serie de consideraciones partiendo de la idea de la indisolubilidad entre democracia y liberalismo, donde el debilitamiento de alguno de sus supuestos significa el adelgazamiento o el quiebre del otro. Coincide en que el componente económico es la causa principal de la debilidad democrática occidental; argumentando que los ciudadanos disfrutaron de un rápido incremento de su nivel de vida después de la gran depresión y hasta la década de los 80, donde por ejemplo, de 1935 a 1960 la renta de un hogar estadounidense típico se duplicó y de 1960 a 1985 volvió a duplicarse.

No obstante, desde esa fecha el ingreso medio norteamericano y europeo se estancó, creando una idea de suma cero, en la que cualquier eventualidad económica corre a cuenta de la clase media.

En un segundo término coloca el componente racial o étnico, que ante el congelamiento del ascenso económico de las clases medias se ha traducido en un temor a perder jerarquía social, lo que está generando una rebelión occidental contra el pluralismo étnico y cultural.

Como un tercer factor de la crisis democrática y solo de una manera referencial, incluye al ascenso de internet como uno de los factores de inestabilidad al separar la distancia entre centro y periferia[9].

De las diversas posturas se desprende que la preocupación sobre la enfermedad democrática resulta inversamente proporcional a la coincidencia que pudiera haber sobre los puntos centrales de su debilitamiento, generando con ello una multiplicidad de dolencias, posiblemente todas reales, pero que dificultan el avance del debate democrático provocando un horizonte de discusión de amplio espectro que lleva a la toma de soluciones parciales que no han ayudado mucho a la mejora de su operación.

Al mismo tiempo, lo que se aprecia de lo anterior es una serie de dolencias atomizadas, propias de una sociedad occidental de la demanda, que en una línea de peticiones y no de causas, la alejan de sus posibles soluciones.

Crece la sensación, dice Lassalle, de que la democracia liberal es un modelo fallido, que los estados de excepción no han restaurado la seguridad, y que el capitalismo cognitivo no ha traído el bienestar para todos, en el marco de una sociedad agotada y resignada ante la adversidad. Que el viejo dilema seguridad o libertad se agudiza y se renueva ahora con la ecuación orden o caos; donde la democracia liberal ha perdido fortaleza ante las profundidades posmodernas de una sociedad ingobernable[10].

9 Mounk, Yascha (2018), *El pueblo contra la democracia*, Espasa, pp.22-24.

10 Lassalle, José María (2021), *El liberalismo herido*, Arpa, pp.12-13.

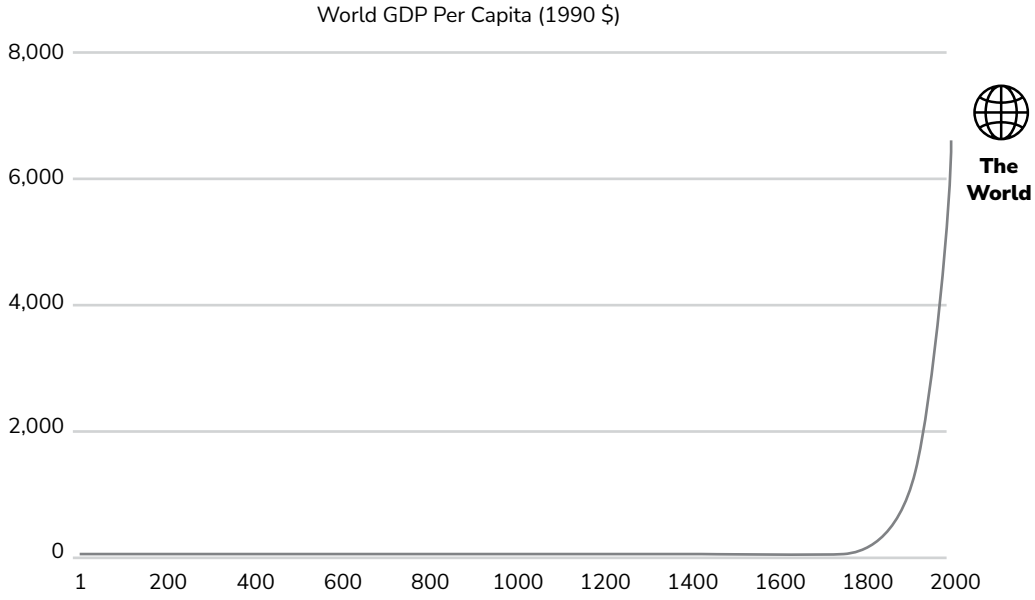
Se desborda la preocupación por resolver un presente democrático que no funciona bien para la mayoría de las sociedades occidentales, respecto a una serie de dolencias que en mayor o menor medida se han venido acumulando desde el nacimiento del liberalismo; olvidando llevar el debate al corazón de un orden liberal que merece una revisión exhaustiva.

III. ¿EL LIBERALISMO AGOTADO?

La parte final del siglo XVII y el siglo XVIII significan un momento estelar en la historia occidental, donde aparecen de manera simultánea una cosmogonía de naturaleza liberal, junto a una Primera Revolución Industrial donde juntas cambiaron la vida del ser humano para siempre.

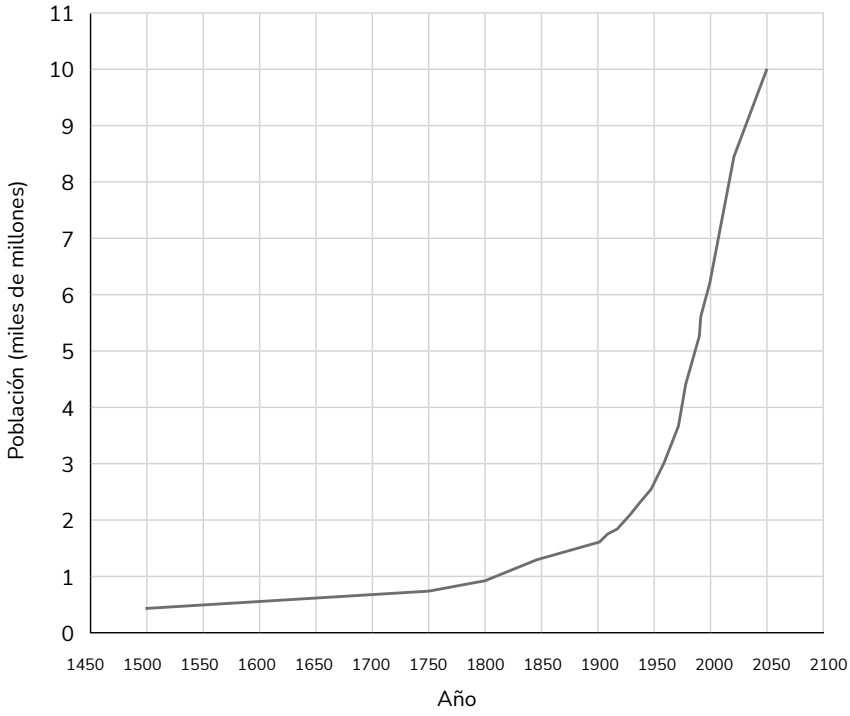
Antes del siglo XVIII, a pesar de las epopeyas premodernas, la humanidad había sido una larga planicie donde lo que sucedía no tenía un impacto general en la población mundial.

GRÁFICA2 PIB Per Cápita Mundial (0-2000)



FUENTE: "Statistics on World Population, GDP, and Per Capita GDP, 1-2008 AD", Angus Maddison; IMF

GRÁFICA3 Evolución de la población mundial (1950-2050)



FUENTE: Research Gate

Es la revolución industrial la que permite el nacimiento del comercio mundial, la manufactura, los incrementos de competitividad, la multiplicación de la innovación, el crecimiento de la economía del mundo, y de manera especial, el descubrimiento de un potencial científico y tecnológico que maravilló a través de sus invenciones al mundo de su época.

Por otro lado, la aparición de un pensamiento lúcido e ilustrado encabezado por Locke, Hobbes, Rousseau, Smith, entre muchos otros, crearon una reflexión que a través de sus propuestas de contrato social, libertad, Estado de Derecho, derechos humanos, mercado, entre otros; tuvo la virtud de conducir y arropar a una innovación industrial y tecnológica que en el liberalismo encontró una explicación de vida que le dio sentido y rumbo.

El liberalismo del siglo XVIII vino a ser una suma de las principales líneas de pensamiento que desde los tiempos griegos intentaron la dignificación del ser humano en sociedad. Sus propuestas sobre la creación de un Estado moderno, de una comunidad social basada en el derecho y en la libertad, dieron lugar a una teoría general del desarrollo que se fue imponiendo en la mayoría de los países occidentales bajo sus propias condiciones.

A lo largo de estos últimos siglos de liberalismo, han surgido diferentes tendencias: liberalismo social, libertario, neoliberalismo, etc. Sin embargo, el liberalismo clásico, agrega Fukuyama, es un gran paraguas bajo el que se cobija una amplia gama de posicionamientos políticos que coinciden en cuanto a la importancia fundamental de la igualdad de los derechos individuales, la ley y la libertad[11].

A la innovación tecnológica industrial, desde su nacimiento, el liberalismo la comprometió con la utopía de una sociedad libre merecedora de un progreso para la mayoría de sus integrantes a través de la organización de un Estado de Derecho que se organiza a partir de la voluntad de la mayoría en una democracia liberal.

La poderosa fuerza del pensamiento liberal, fue una piedra de toque de un liberalismo económico que se convirtió en una constante para la mayoría de las naciones occidentales de su tiempo, las cuales fueron instrumentando un capitalismo manufacturero ante la aparición de un poderoso sector secundario que desplazaba rápidamente a la economía agrícola.

Este liberalismo económico, ante sus propias contradicciones y excesos, vivió con intensidad los ajustes sociales que fueron apareciendo en la segunda parte del siglo XIX y que explotaron en la primera mitad del siglo XX.

Después de esta convulsiva etapa de reacomodo político, económico y social, a partir de 1950 vivió una segunda etapa muy exitosa donde el comercio global tuvo tasas de crecimiento aproximadas de 7.4% anual promedio y la economía en su conjunto de un 5% anual promedio hasta mediados de la década de los 70 (Ocampo, 2017), donde el mundo pudo demostrarse a sí mismo que podían convivir de manera armoniosa tanto el comercio global como el desarrollo económico, y de manera importante, la mejora de las condiciones sociales de las personas que hacían posible la marcha de la industria y la manufactura en el mundo.

No obstante, las crisis presentadas en la década de los 70 detuvieron el avance de estos logros y en el marco de los conflictos del petróleo, del aumento de las materias primas y los alimentos, entre otros, abrieron la puerta para que se introdujeran corrientes económicas como el neoliberalismo, que en congruencia a la línea de sus postulados, expulsó al Estado de su responsabilidad de orquestador del desarrollo económico, condenándolo al papel de testigo de los hechos económicos.

---

11 Oropeza, Arturo. (2024), *Hacia una nueva visión del desarrollo industrial digital en México*, p.39.

Margaret Thatcher en Inglaterra, Ronald Reagan en Estados Unidos, y otros en la mayoría de los países occidentales y su periferia, bajo diversas adecuaciones fueron orientando sus estrategias económicas hacia las líneas neoliberales que dictaba un Consenso de Washington, que como una receta milagrosa la ofrecía para salir de la trampa del sub-desarrollo y llegar al nivel de ingreso medio de los países desarrollados.

Desde finales de la década de 1970, reconoce Fukuyama, el liberalismo económico evolucionó hacia lo que actualmente se denomina neoliberalismo, el cual incrementó drásticamente la desigualdad económica y provocó devastadoras crisis financieras que han perjudicado a la gente corriente mucho más que a las élites adineradas en muchos países del mundo. En esta desigualdad, agrega el mismo autor, se basa el argumento progresista en contra del liberalismo y sistema capitalista asociado. Esta desigualdad, este desencanto sobre el liberalismo económico ha evolucionado en las últimas décadas, generando demandas tanto de derecha como de izquierda, pidiendo que esta doctrina sea sustituida de raíz por un sistema diferente[12].

A pesar de existir una idea generalizada del fracaso del neoliberalismo, el debate de su procedencia continúa en un reciclaje de filias y fobias con relación al papel que el Estado debe tener en el desarrollo económico.

Aparecen en este nuevo debate los sucesores de Thatcher y de Regan, como Meloni de Italia, Milei en Argentina, etc., quien en su último viaje a Europa declaró que "La justicia social es aberrante", "el control de capitales es inmoral" y "los impuestos a empresarios son un robo"(El País, 18 de mayo 2024).

Frente a autores que reconocen el fracaso del neoliberalismo económico como Mazzucato, Rodrik, el propio Fukuyama, etcétera, aparecen autores como Easterly, Couchrane, McCloskey, entre otros, que siguen defendiendo a la corriente neoliberal. McCloskey por ejemplo, al propio tiempo que critica las ideas de Mazzucato sobre un Estado emprendedor, agrega que los gobiernos tienden a tomar decisiones estúpidas y que la innovación genuina surge mejor en un entorno de libre mercado y que los gobiernos no deben asumir un papel central en dirigir la economía. Easterly reitera estas posturas invocando a Adam Smith, señalando que el desarrollo económico se basa en el libre comercio y el libre mercado.

Los costos de los errores cometidos en esta batalla por el desarrollo, poco a poco se han ido reconociendo por los líderes políticos occidentales. En el caso de Macron (Francia) al declarar que "La desindustrialización francesa es una de

12 *Ibidem*., pp.40-41.

las causas de nuestra desgracia". En el caso de Estados Unidos cuando Biden argumenta sobre el olvido industrial del país[13]. En el caso de Trump, cuando de manera paranoica arremete contra el mundo para recuperar un *supuesto desarrollo robado*.

Al fracaso económico que se reconoce del modelo neoliberal ante la insatisfacción de sus clases media y populares; pero también de su competencia geopolítica y económica comprometida con el acenso exitoso de China y la región de Asia del Este; habría que añadir ahora la aparición de un nuevo capitalismo digital que lo compromete con sus líneas esenciales.

Después de casi tres siglos de haberse presentado como un relato virtuoso y convincente de la cosmogonía occidental, el liberalismo a través de sus diferentes manifestaciones y escuelas, se enfrenta ahora a uno de sus retos más importantes, que va desde la posibilidad de replantearse a través de sus conceptos y valores originales, hasta el punto de su propia desaparición.

Los liberales, señala Noah, a la fecha no comprenden como la historia se desvió de su ruta predestinada, y carecen de un prisma alternativo que les ayude a interpretar la realidad[14].

En una visión que no duda en dibujar escenarios complejos para la sociedad liberal, Innerarity argumenta que entendemos muy poco nuestro tiempo. Que no estamos comprendiendo cómo funciona esta sociedad y cuáles son nuestras posibilidades de acción con ella. Que el simplismo de algunos de los análisis procede de la falta de actualización de los conceptos políticos que fueron pensados en una época de relativa simplicidad social y política, antes de los grandes conflictos sociales que inauguraron el mundo contemporáneo con sociedades que no conocían el actual pluralismo cultural, político y tecnológico[15].

La democracia y el liberalismo, categorías que nacen juntas en el tiempo, también enfrentan de manera conjunta ese cambio político, social, económico y tecnológico que se les vino sin presentir y que a las sociedades occidentales las encuentra desprovistas del instrumental que les permita proponer las medidas de solución para preservar los elementos esenciales del liberalismo como la propia libertad, desarrollo económico y seguridad.

La creencia del fin de la historia soportada con la caída del relato comunista soviético de 1991, derivó en la complacencia de creer que el liberalismo se alzaría

13 *Ibidem*., p.42.

14 Noah, Yuval (2018), *21 lecciones para el siglo XXI*, Debate, p.23.

15 Innenarity, Daniel (2020), *Una teoría de la democracia compleja Gobernar el siglo XXI*, Gutenberg, p.13.

en todo el mundo como una verdad universal que conduciría a la humanidad a su progreso y estabilidad.

Se olvidó que el liberalismo, como todo relato producto de la construcción social, depende de las políticas de remedio que oportunamente se le puedan inyectar para su subsistencia.

La aparición precipitada de su vertiente neoliberal en la década de los 80, destruyó parte de sus equilibrios y postulados, ocasionando un aumento en las disparidades económicas de los diferentes segmentos sociales; tanto de los países centrales como periféricos.

Al respecto Zuboff nos habla de la colisión que se ha producido entre los procesos históricos de individualización que conforman nuestra experiencia como seres autónomos, tras décadas de vigencia de un régimen de economía de mercado neoliberal, en la que la conciencia de nuestra propia valía personal y nuestras necesidades de libre determinación individual autónoma se han visto sistemáticamente boicoteadas por el otro[16].

Para los efectos de este ensayo la parte a destacar se refiere al fracaso de un neoliberalismo que alejó al Estado de su responsabilidad de ser el orquestador del desarrollo, llevándolo a una disminución de su poder, donde su no regulación y orientación de lo económico fueron tomadas por el mercado, en detrimento de los pesos y contrapesos que siempre se pensaron para un Estado de Derecho.

A esta debilidad del liberalismo occidental habría que sumar la llegada intempestiva de una Primera Revolución Digital, que advertida con toda su fuerza a partir de la primera década del siglo XXI, ha demostrado con su llegada, como lo señala Noah, que la sociedad occidental no estaba preparada para administrar un cambio de estas dimensiones, el cual está poniendo al liberalismo contra la pared de su sobrevivencia.

En la línea dogmática del propio neoliberalismo que se ha desarrollado el último medio siglo, donde prevaleció la idea de un mercado poderoso y un estado débil, la aparición de un *nuevo mercado de los datos* liderado por las *Big Tech*, lo que, ha encontrado en su alumbramiento es un entorno de total permisividad sin los marcos adecuados de regulación. Y tal vez de manera más importante, sin que se hayan formulado como lo hicieron los filósofos liberales de los siglos XVII y XVIII, las preguntas primigenias de ¿qué queremos como sociedad con la nueva civilización digital?, y de ¿cuál es el lugar que queremos darle al ser humano a lo largo del presente siglo y tiempos por venir?

16 Zuboff, Shoshana (2020), *La era del capitalismo de la vigilancia*, Paidós, pp.34-35.

Lassalle comenta que "Bajo este panorama la sociedad se ha hecho ingobernable, pues el imperativo neoliberal asumido por el capitalismo cognitivo de las grandes plataformas y los datos disuelve más y más las bases consensuales de racionalidad cívica y contractualista del liberalismo de estirpe lockeana y nos aboca a una forma pura de dominación algorítmica sin contestación ni disidencia"[17].

IV. LIBERALISMO, DEMOCRACIA Y ALGORITMOS

La sociedad global en su conjunto, auspiciada en la lógica de sus propios relatos, enfrenta desde finales del siglo pasado una disrupción civilizatoria provocada por el avance tecnológico de una Primera Revolución Digital que como su antecesora, la Primera Revolución Industrial, viene a retar todo lo conocido; cuestionando a cada una de sus estructuras y poniendo una vez más al hombre frente a su sentido de futuro.

El avasallamiento que hoy provoca un sinnúmero de tecnologías digitales que están cambiando la vida del ser humano con una velocidad vertiginosa, solo es comparable con la conmoción que causó en su momento a la sociedad de 1750 la aparición de la innovación industrial.

En Occidente estamos ante el surgimiento de una nueva era digital, de un nuevo poder digital, el cual a través de la fuerza del algoritmo amenaza con maravillar e impulsar el avance civilizatorio de la humanidad, de igual modo que de controlar su destino.

La Revolución Digital es muchas cosas. Es una transformación digital de lo intangible, de los algoritmos y los genes. De la llegada de una nueva tecnología 5G y sus aceleradores de velocidad; de las latencias en milisegundos; del aumento de la tasa de datos; de los altos gigabytes por segundo; de las bandas anchas miles de veces más rápidas; etcétera; que traducidas en aceleradores de la innovación se desbordan en cómputo en la nube, centros de datos, big data, internet de las cosas, robótica, ciberseguridad, biotecnología, etcétera, y como la cúspide de este geométrico desarrollo, una inteligencia artificial que emplaza desde ahora la evolución y destrucción de su propio creador.

La inteligencia artificial como nombre y concepto no es reciente. Ya en 1956 John Mccarthy acuñó el nombre y desarrolló el primer lenguaje de programación de inteligencia artificial. En 1950 se fabricó el primer ordenador

17 Lassalle, José María (2021), Op. Cit, p.22.

que permitió el juego de ajedrez entre humano y máquina; en 64-66 *Eliza* fue el primer Chatbot que simuló una conversación entre humanos y máquinas; en 1997 la supercomputadora *Deep Blue* derrotó al campeón mundial de ajedrez Gary Kaspárov, como una premonición de lo que está siendo ya la inteligencia artificial; y en 2016 *Alpha Go* derrotó al campeón mundial de la especialidad con jugadas imprevisibles que denotaron el éxito de los sistemas de inteligencia digital (Vanguardia 2020).

En materia de inversiones, la inteligencia artificial alcanzará los 200 mil millones de dólares para 2025 y se espera escale a 2 billones de dólares para 2030; estimando que solo los sistemas y aplicaciones de I.A. generativa producirán 7 billones de dólares en 2030[18].

La fascinación y el espanto que causa hoy la inteligencia artificial ha recorrido un largo camino desde sus orígenes militares, donde el pentágono desde finales de la Segunda Guerra Mundial ha estado apoyando económica y tecnológicamente a la mayor parte de la innovación que está en el mercado[19].

La inteligencia artificial llega a una primera etapa de madurez, gracias su mejora técnica y a la acumulación de datos que se han estado produciendo en las redes sociales, los cuales le han permitido escalar en información y poder, provocando una división de la sociedad entre aquellos que de manera eufórica encuentran apasionante el desenlace de un futuro digital que desconocen, y una sociedad ludita temerosa de no saber que sigue.

Los dueños de las nuevas armas del poder digital como Mark Zuckerberg, dueño de Meta, defienden que cualesquiera que sea la nueva tecnología que venga a cambiar la naturaleza de las cosas, siempre hay gente que se preocupa de estos cambios y desea regresar al pasado. Asimismo, empresarios del sector indican que la mayoría de los temores sobre la inteligencia artificial están sobrevalorados, si no es que totalmente infundados. Que solo será hasta que la tecnología esté integrada en su totalidad a la vida de las personas cuando paren de temerle[20].

Otros como Lassalle, encabezando la línea radical de la corriente que advierte de una enorme amenaza para la humanidad ante el manejo de una inteligencia artificial sin regulación, indica:

18 López-Portillo, José Ramón; Blog Nexos, julio 2024, p.29.

19 Las agencias norteamericanas de inteligencia artificial, señala Crawford, han sido los viejos guardianes de la big data, y a través de los grandes proyectos de investigación de defensa, han sido los principales conductores de la inteligencia artificial desde los inicios de los cincuenta. (Crawford, Kate (2021), *Atlas of AI*, Yale. p.184).

20 Acemoglu, Daron y Johnson, Simon (2023), *Power and Progress*, Publicaffairs, pag. 373.

"Una figura destaca sobre el horizonte de incertidumbres, malestares y miedos que acompaña el comienzo del siglo XXI. Se trata por ahora de una silueta por definir. Una imagen que todavía no refleja con exactitud sus contornos pero que proyecta una inquietud en el ambiente que nos previene frente a ella. Su aparición delata un movimiento de alzada vigorosa, que lo eleva sobre la superficie de los acontecimientos que nos acompañan a lo largo del tránsito del nuevo milenio".

"Envuelta por un aliento de energía sin límites, su forma va adquiriendo volúmenes titánicos en los que se presiente la desnudez granítica de una nueva expresión de poder. Con sus gestos se anuncia el reinado político de un mundo desprovisto de ciudadanía, sin derechos ni libertad. Una época que asistirá a la extinción de la democracia liberal. Que instaurará una era mítica a la manera de las que imaginó Hesíodo, hecha de vigilancia y silicio, habitada por una raza de humanos sometidos al orden y a la seguridad. Un mundo de fibra óptica y tecnología 5G, dominado por una visión poshumana, que desbordará y marginará el concepto que hemos tenido del hombre desde la Grecia clásica a nuestros días"[21].

Geeks y luditas; tecnófilos y tecnófobos, se dividen ante la llegada de una IA donde los primeros la ven como un santo grial que resolverá todos los problemas del mundo; mientras que el otro grupo la aprecia como un enorme reto que como ya se señaló, atenta no solo al sistema liberal y democrático, sino que ataca el trabajo y la forma de vida del ser humano, condenándolo a la intrascendencia.

En este punto valdría la pena recordar a Suleyman cuando advierte de no caer en la *trampa de la aversión al pesimismo*, que surge cuando uno se siente abrumado por el miedo a enfrentarse a realidades que podrían ser nefastas, y a pesar de ello mira hacia otro lado[22].

Es en este entorno de los asombros y los miedos donde se ha ido colando sin mucho aspaviento la amenaza más grande que se ha intentado contra la libertad del ser humano y por ende de sus instituciones como la democracia y el propio liberalismo.

Amparado en un debate democrático infinito, la realidad digital avanza más rápido de lo que se supone, introduciéndose silenciosamente en la voluntad de

21 Lassalle, José María (2019), *Ciberleviatán*, Arpa, pp. 17-18.

22 Suleyman, Mustafa (2024), *La ola que viene*, Debate, p. 22.

la sociedad actual mientras esta discute alrededor de una agenda interminable donde aparecen todo tipo de temas de naturaleza económica, política y social.

Se abren seminarios y múltiples debates sobre si restaurando la economía se podrá mejorar la democracia; si modificando una vez más el barroquismo de sus formas se le podrá mejorar; si incluyendo o no los temas de un sin número de minorías arribaremos a un mejor estadio de libertades políticas, etc., mientras oportunistas políticos avezados, advirtiendo de la anarquía del momento, ofreciendo una oferta infinita de mejoras que no podrán cumplir, van construyendo diversas formas de populismo tanto de izquierda como de derecha, que en una primera instancia los presenta como los ganadores del desconcierto.

Sin embargo, más allá de la procedencia de cada uno de los derechos que se intentan reivindicar para contar con una democracia mejor; de manera paralela ha estado avanzando una humedad que se cuela poco a poco en cientos de millones de seres humanos; un mundo panóptico disfrazado de utopía que va mermando progresivamente tanto los ámbitos de libertad del ser humano como la fortaleza del Estado occidental.

V. HACIA UNA UTOPÍA DE LO DIGITAL

El inicio de la construcción del poder digital en su versión política es reciente y se da junto con la evolución mostrada por el avance de su propio desarrollo a lo largo de este siglo.

Si bien la aparición de Microsoft (1975), Apple (1976), Nvidia (1993), Amazon (1994), Google (1998), Alibaba (1999), Tesla (2003), Facebook (2004), etc., vienen desde el último tercio del siglo pasado, la velocidad de sus cambios las ha llevado rápidamente a desdoblarse en nuevas tecnologías que cambiaron la lógica y el sentido de sus negocios.

Para que la revolución industrial pasara de su primera etapa a lo que se llamó la segunda revolución industrial, a la cual se le ubica con el nacimiento del fordismo, tuvieron que pasar 150 años. En el caso de la revolución digital, el disparo de la inteligencia artificial, como epítome de su innovación, se ha dado en un espacio aproximado de seis décadas.

La tiranía de la Ley de Moore y la contundencia de la Ley de Koomey en el siglo XX, que se han venido comprobado en los hechos cada 18 meses, que indican que la humanidad bajaría los costos digitales al propio tiempo que aumentarían sus potenciales de velocidad y almacenamiento; junto con la Ley de Kurzweil de rendimientos acelerados, que vaticina que la inteligencia artificial a 2050 se desarrollará de tal forma que la capacidad de cómputo de

los ordenadores excederá por más de mil veces la de un cerebro humano, nos ha llevado a la tercera década del siglo a vivir escenarios que nunca nos hubiéramos imaginado.

La velocidad de estos resultados ha sorprendido a todos. Desde los dueños de las propias empresas tecnológicas, a los diversos analistas y desde luego a un Estado descuidado que preocupado por un relato y una problemática propia del siglo pasado, no acaba de enterarse de las nuevas estructuras de un cambio civilizatorio que amenaza con transformar lo conocido.

En el caso de Estados Unidos, desde siempre líder del avance digital como resultado de los avances científicos y tecnológicos de postguerra, consciente de su importancia en la confrontación geopolítica y económica, hasta la fecha ha mantenido el monopolio de la innovación digital en asociación estratégica con su sector privado. China, Rusia, Israel y otros países también han logrado avances significativos en lo que Crawford denomina como *la guerra de la inteligencia artificial*.

Sin embargo, el éxito obtenido por estas empresas tecnológicas ha sido de tal magnitud, que su poder las ha llevado a constituirse en verdaderos gigantes económicos por capitalización bursátil, donde Apple ($3,33 billones de USD), Microsoft ($3,11 billones de USD), NVIDIA ($2,81 billones de USD), Alphabet ($2,03 billones de USD) y Amazon ($1,88 billones de USD), se presenten hoy como las cinco empresas más grandes del mundo[23].

Este nuevo poder económico digital ha llevado a que estas empresas sean más poderosas económicamente que países como España, Brasil, Australia, Rusia, etc., a los cuales cada una de ellas supera en valor económico.

El negocio digital nace fuertemente a fines del siglo XX, siendo la velocidad de sus dividendos lo que les ha redituado un nuevo poder económico-político. Estos dividendos surgen de un poder de la innovación digital que las coloca en un plano de monopolios respecto a otros actores económicos en el mundo. En el caso de Windows, por ejemplo, controla el 73% del mercado mundial de sistemas operativos para ordenadores; en el tema de sistemas operativos para los teléfonos celulares, Android también domina 73% del mercado mundial; o en el empleo de nubes de almacenamiento de datos, en lo que Amazon (AWS), Microsoft (Azure) y Google (Cloud), controlan el 70% del mercado mundial[24], etc.

El negocio del internet por su propia naturaleza nace exitoso. Desde su inicio fue recibido con grandes expectativas y sus novedades tecnológicas han sido

23 State Counter, El país, 27 de julio 2024.

24 Market Cap. El país, 27 de julio 2024.

celebradas por la sociedad global en su conjunto. Sin embargo, el cambio que determina un antes y un después del negocio digital es la enorme compilación de datos que se fue dando como resultado natural del uso de los diferentes aparatos de tecnología, que a 2021 ya acumulaban un número estimado de 1900 billones de sitios(websites)*.

La aparición de las primeras redes sociales llevó a la creencia, como lo llegó a señalar Hillary Clinton en 2010, que estas eran los nuevos pilares de la democracia mundial. Uno de los líderes de las protestas egipcias de la llamada Primavera Árabe, Wael Ghonim, también comentó que a él le gustaría entrevistarse con Mark Zuckerberg para agradecerle personalmente la participación de las redes en estos movimientos, los cuales en su mayoría habían comenzado a través de Facebook; agregando que si se quería liberar a las sociedades oprimidas, justo se les debería de proporcionar internet[26].

No obstante, esta primera idea de libertad multiplicada a través de la internet como instrumento para la liberación, poco a poco ha ido cambiando ante otro tipo de experiencias que también se han presentado en el horizonte de la tecnología digital.

El caso de Myanmar de 2010- 2017, como uno de los primeros eventos donde la desinformación vertida por internet exacerbó un sentimiento racial en contra de una pequeña secta musulmana de los Rohingya, llevó a un problema social que se catalogó de genocidio. En este evento pudo evidenciarse que esta tragedia había sido fomentada por las redes sociales y que incluso dentro de ellas se organizaron los crímenes a través de la plataforma de Facebook. Ante los hechos, la plataforma lo único que se atrevió a comentar sobre la masacre es que estaba de acuerdo de que debió haber hecho más por evitarla[27].

Estos ejemplos que empezaron a mostrar la otra cara del *fenómeno tecnológico*, fueron expandiéndose ante el crecimiento y poder de las plataformas sociales, que rápidamente se alejaban de sus potenciales de libertad en favor de la democracia, y se acercaban más a un mundo de libertades dirigidas o asistidas que iba surgiendo del poder de las propias plataformas.

---

25 En cuanto a la inteligencia artificial, esta inicia su entrenamiento con datos etiquetados a mano, como el caso de ChatGPT 2, con 1500 millones de parámetros. Gracias a la acumulación de datos, a 2023, sin etiquetar los datos a mano, la aplicación Inflection AI, ya contaba con 540 mil millones de parámetros. En el caso de Alibaba, se habla de que cuenta ya con 10 billones de parámetros.

26 Acemoglu, Daron y Johnson, Simon (2023), Op. Cit, p. 342.

27 *Ibidem*, pp. 356 y 357.

En 2018 Sri Lanka continuó con el antecedente de Myanmar a través de un discurso de odio en contra de los musulmanes. En 2020 en India también se dieron eventos de asesinatos y destrucción de mezquitas en contra de la secta Rohingya musulmana.

El mundo pudo tener una idea más clara de lo que empezaba a representar el poder digital cuando la empresa Cambridge Analítica reconoció públicamente junto con Mark Zuckerberg, dueño de Facebook, que había comprado de manera indebida la información de más de 50 millones de usuarios de la red social de Estados Unidos para objetivos políticos y comerciales, entre ellos las elecciones donde salió como ganador el presidente Trump.

Desde ese momento quedó claro que se abría un mundo de la vigilancia de los datos sociales, que habían encontrado un medio idóneo para difundir hechos falsos de una manera más rápida, profunda y general, que lo que lo hacía la propia verdad en las diferentes categorías de la información. También se pudo establecer que las redes sociales eran la primera fuente de información en Estados Unidos y que el 70% de lo que la gente leía era enviado de manera dirigida por dichas plataformas. Que de 100 publicaciones o recomendaciones que se hacían en Twitter, Facebook, etcétera, la mayoría contenía desinformación o teorías de conspiración[28].

No solo fue la imparable velocidad de reunir cientos de millones de usuarios en las redes sociales lo que empezó a abrir el apetito de las grandes empresas de tecnología, que empezaron a ver cómo sus millones se convertían en billones de dólares con gran velocidad. El impacto fue constatar cómo en tan poco tiempo se habían hecho de un enorme archivo de datos personales que no habían contemplado y que ahora representaban un gran poder de información para cada una de ellas.

En ese sentido el paso del internet a la datología y luego a la inteligencia artificial fue una dialéctica de la información y del lucro, que ahora se ha transformado en un poder político-económico de grandes dimensiones para las empresas Big Tech.

Larry Page y Sergey Brin fueron unos de los principales especialistas que encontraron la mejor combinación para sumar la tecnología con el dato humano, al crear los algoritmos que pudieron vincular exitosamente las dos categorías. En última instancia, como señala Acemoglu, cómo el algoritmo calcula los resultados es secundario, el avance importante es que Page y Brin idearon una

28 *Ibidem*., p. 359.

forma de utilizar los conocimientos e ideas humanas al mejorar una tarea clave de la máquina: clasificar los resultados de búsqueda[29].

La monetización de los datos, el nacimiento de un nuevo modelo de negocio, la posibilidad de construir un verdadero poder digital, surge en este momento en el que la innovación logarítmica fue tomada por Google para mercantilizar esta nueva tecnología de los datos a través del importante concepto de los **anuncios digitales dirigidos**, lo cual vino a revolucionar el propio avance y tendencias que hasta ese momento registraba la primera revolución digital.

Por ello en 1998 e incluso 2000, como comenta Acemoglu, casi nadie tenía una idea clara acerca de lo que era el big data y la inteligencia artificial. La nueva asociación entre el algoritmo y el dato, traducida en la expansión de los anuncios dirigidos, vino a generar una explosión autogenerativa de datos y más datos que han servido de laboratorio para el perfeccionamiento de una inteligencia artificial generativa, que técnicamente puede asimilar la historia de la humanidad en medio día y puede estar hoy en posibilidad de generar lo imposible.

Agrega Acemoglu que nosotros no podríamos entender el desequilibrio político que han venido a causar las redes sociales, si no agregamos el desmedido incentivo económico basado en los anuncios dirigidos, los cuales derivaron del potencial digital de optimizar los gustos, los rangos, las caras y ahora incluso los futuros de la sociedad global[30].

Todo este acontecimiento que envuelve al fenómeno de la datología digital está derivando, como lo indica Zuboff, en un nuevo capitalismo de la vigilancia que emerge del poder que están acumulando con gran rapidez las Big Tech[31], las cuales sobrepasando las ambiciones tradicionales del capitalismo, reclaman ahora el dominio sobre territorios humanos, sociales y políticos que se extienden mucho más allá del ámbito convencional de la empresa privada o del mercado; lo cual se traduce en un derrocamiento de la sobernía del pueblo y de una

29 *Ibidem*, p. 365

30 *Ibidem*, p. 371.

31 Facebook (o Meta, como ahora se conoce) y Google (o Alphabet, como ahora se conoce), tienen poco en términos de instalaciones físicas, pero grandes sistemas y procesos para recolectar, almacenar y procesar datos para conectar a las personas con el conocimiento, el análisis y entre sí. Son monopolios globales naturales, no en el sentido multinacional tradicional de tener grandes plantas físicas dispersas por todo el mundo, sino en dominar mercados donde los clientes se benefician de estar conectados globalmente. Son sus complejas redes de personas e información, no su infraestructura física, lo que los competidores no pueden replicar ni reemplazar (Mayer, Colin (2024), *Capitalismo y Crisis. Como arreglarlas*, Oxford, p.8).

peligrosa deriva hacia la *desconsolidación* de la democracia que actualmente amenaza a los sitemas liberales de Occidente[32].

La vigilancia, la mentira, la desinformación, la expropiación de la voluntad, la inducción del deseo económico, político y social, el desempleo digital[33], son las puertas que se abren al infinito ante este nuevo impulso del cambio digital.

32 Zuboff, Shoshana (2020), Op. Cit. p. 38.

33 *"Antes o después, comenta Skidelsky, nos quedaremos sin empleos" (Carr, Nicholas, *Atrapados. Cómo las máquinas se apoderan de nuestras vidas*, 2014, p.47). El tema no es nuevo, nace junto con la amenaza generalizada del desempleo que surgió a partir de la Primera Revolución Industrial, cuando las primeras máquinas vinieron a sustituir mucho del trabajo físico del campo y los talleres; al mismo tiempo que se crearon millones de empleos en la nueva industria de la manufactura y sus servicios. En el siglo XIX David Ricardo (1817) preveía que una mejor distribución del trabajo a través de las máquinas desplazaría la actividad humana hacía la ciencia y el arte. Douglas Haberler y otros, reconocían la posibilidad del desempleo digital pero agregaban que sería compensado automáticamente. (Acemoglu, Daron, (2023), Op. Cit. pp. 434-435). En 1930 Keynes advertía que "estabamos afligidos" por una nueva enfermedad de la que algunos lectores pueden no haber oído el nombre, pero de la que oirán mucho en los años venideros, a saber, "**el desempleo tecnológico**". Por otro lado, Difazio, Rifkin, Brynjolfsson, McAfee, Spence, etc., en las últimas décadas han venido retomando el tema con nuevos argumentos. McAfee, por ejemplo, demuestra como la tecnología ha venido aumentando velozmente la productividad de Estados Unidos, junto con beneficios empresariales, mientras que el número total de empleos en el país apenas ha crecido. Spence, agrega "La sustitución de empleos manuales tradicionales por máquinas y robots es una tendencia fuerte, permanente y que quizás esté acelerando la fabricación y la logística, mientras redes de computadoras reemplazan empleos administrativos en el procesamiento de información". Sobre el tema Krugman agrega que las máquinas inteligentes podrán aumentar el PIB, pero que al mismo tiempo reducirán la demanda de gente, incluida la gente inteligente; en la que la mayor parte de las ganancias serán para los dueños de la robots, o sea las Big Tech. (Carr, Nicholas, p.40-45). Rifkin, en su libro clásico de "El fin del trabajo", además de denunciar un desempleo mundial de 800 millones de seres humanos, añadia que "millones de trabajadores ya han sido eliminados de los procesos económicos y los diferentes tipo específicos de empleo han sido reducidos en importancia o bien reestructurados o han desaparecido". (Rifkin, Jeremy (1996), *El fin del trabajo*, Grupo Planeta, p.17).

VI. DE DATOLOGÍAS Y DISTOPÍAS

Desde el *zoon politikón* de Aristóteles, cada una de las sociedades de su tiempo partieron de la consideración de un ser humano como centro de una sociedad que de diferentes maneras debía construirse para su bienestar y felicidad.

Es cierto que en muchas etapas de su historia el ser humano ha fracasado en lograr estos objetivos, pero su anhelo de libertad y de una vida mejor siempre han estado presentes a lo largo de su camino.

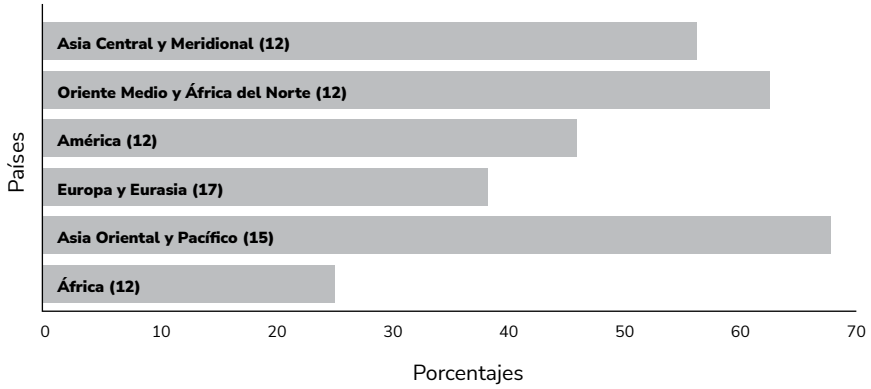
En esta última etapa, mezcla de liberalismo y revolución industrial, la reiterada búsqueda de la libertad y del bienestar han sido los principales motores en el constructo político de un Estado de Derecho definido a través de una participación democrática; en la idea de una sociedad industrial que sintió que el progreso que había llegado podía ser para siempre.

Esta utopía de libertad y progreso enfrenta un momento crucial en la que una nueva cosmogonía sin autores, pero con el sólido argumento de los hechos, lo que busca por primera vez en la historia de la humanidad es penetrar en el ámbito de libertad del ser humano, cualesquiera que este haya sido, e influir poco a poco en sus ideas, en sus gustos, diversiones, preferencias, valores, trabajos y visión de futuro, a través de una libertad asistida que poco a poco le viene a decir cuál debe ser su comportamiento y cuál deberá ser su papel en la nueva sociedad de la vigilancia.

En este sentido pareciera que el papel del ser humano empieza a reducirse en primer lugar a una ecuación de usuario-cliente, donde a través de sus propios datos, de su conducta social, política y económica, se le vigila, se le desinforma, se le miente, se le manipula, se le desemplea, se provoca su cambio de conducta, de sus preferencias y al final se le coloca en un estado de incertidumbre y desculturización que lo arriesga a un estado de robotización o ansiedad; o peor aún, de intrascendencia.

En un segundo plano, de un ser humano reconocido en sus valores y defectos, se le pasa a considerar una materia prima del nuevo mecanismo de la producción digital; que lo oferta al mercado como una mercancía más en busca de utilidades económicas, comerciales, políticas y sociales.

GRÁFICA 4 Porcentaje de países en cada región que han adoptan la vigilancia mediante IA



NOTA Los números entre paréntesis indican cuántos países han adoptado la vigilancia mediante IA en la región.
FUENTE: Vanguardia 2020

En la síntesis de esta ecuación, lo que se genera es un nuevo tipo de poder tanto económico, como político y social, donde en los mercados de la datología humana los compradores adquieren perfiles, fotos, conductas, audios, pensamientos, creencias, gustos, etcétera, que como los nuevos insumos de un nuevo proyecto de producción digital, se están usando ya para aumentar las ventas de productos, incrementar las utilidades de las empresas; para moldear conductas sociales y políticas, y peor aún, para inhibir las expectativas políticas que se fueron creando a lo largo de siglos de civilización.

Esta nueva realidad parasitaria que se alimenta de la conducta del ser humano, amenaza de manera directa la existencia o el poder del propio Estado y consecuentemente de su democracia y su libertad, ante la expropiación de la voluntad humana.

Zuboff comenta que a medida que la competencia se intensifica, los capitalistas de la vigilancia aprenden que no basta con extraer experiencias humanas. Las existencias de materia prima más predictivas de todas son las que proceden de intervenir en nuestra experiencia para moldear nuestro comportamiento, de tal modo que favorezca los resultados comerciales pretendidos por los capitalistas.

Por esta razón, las Big Tech diseñan nuevos algoritmos para influir en la conducta humana y a su vez modificarla, mientras los medios de producción se subordinan a un nuevo y más complejo medio de modificación conductual[34].

34 *Ibidem*, p.36.

Este nuevo poder, continúa Zuboff, aspira a organizar, estimular y afinar a la sociedad para que alcance una confluencia social análoga, en la que la presión del grupo y la certeza computacional sustituyan a la política y a la democracia y así se extingan tanto la realidad sentida como la función social de una existencia individualizada[35].

No somos sus clientes, somos su *producto* dice Noah, en el marco de una carrera por los datos que ya ha empezado de parte de Big Techs como Google, Facebook, Microsoft, Baidu, etc., los cuales primero captan nuestra atención al proporcionarnos de manera gratuita información, servicios y diversión, para después revender nuestra atención a los anunciantes. El verdadero negocio consiste en acumular cantidades inmensas de datos sobre nosotros que valen más que cualquier ingreso publicitario[36].

Estos datos han sido expropiados sin nuestro consentimiento a través de diversos mecanismos que han llegado a un alto nivel de cinismo e impunidad, como se demostró desde su momento en el caso de Cambridge Analítica, donde se aceptó que los datos que se habían confiado a Facebook eran recogidos por terceras partes y usados para manipular las elecciones de diversos lugares del mundo. Estos y otros ejemplos han venido dando fe de la impunidad en el manejo de los datos de las redes sociales.

De manera reciente (2024), solo para enunciar brevemente esta tendencia, se demostró que Elon Musk cambió sin avisar la configuración de datos de su plataforma X para permitir que su nuevo modelo de inteligencia artificial llamado Grok, fuera entrenado con todos los twits enviados a la red sin que los usuarios supieran y sin que hubieran dado su permiso. Después de ello, ha estado utilizando esta misma plataforma para tener entrevistas con el candidato Trump y otros políticos europeos de ultra derecha, en un ejercicio abusivo de las herramientas digitales de su propiedad[37].

Como se aprecia con transparencia a partir de enero de 2025, la alianza del poder político con el poder digital, ha quedado evidenciado a través del binomio Musk-Trump, en un avance sin sonrojos de lo que puede llegar a ser en el corto plazo la infocracia.

35 *Ibidem*, p.37.

36 Noah, Yuval (2018), Op. Cit. 100 y 101.

37 El País, 07 de Agosto, 2024.

Este nuevo capitalismo digital o de la vigilancia se presenta como un proyecto sin escrúpulos que ignora el Estado de Derecho, que se apropia de lo ajeno expropiando los atributos del ser humano, de su individualidad; en la configuración de un nuevo universo digital que empieza a moverse en su propio espacio e interés.

Aunque bajo líneas diferentes, pareciera que este nuevo capitalismo digital se adhiere a los lemas del siglo XIX cuando se señalaba que la *Ciencia descubre. La industria aplica, y el hombre se acopla* [38], en una filosofía a modo que le pide al ser humano su total rendición respecto a los intereses de las Big Tech ¿Por qué esta impunidad? ¿Por qué esta falta de regulación y control sobre un tema trascendente para el ser humano?

VII. REGULAR O NO REGULAR. THATS THE QUESTION?

El neoliberalismo, como ya se adelantó, ha sido el principal campo de cultivo para el arribo de una revolución digital sin límites que se ha aprovechado de la debilidad de un estado occidental que fue adelgazado y recluido a su mínima expresión por una teoría que creía que el mercado abarcaba todo.

En esta idea de la no regulación, sobre todo en Estados Unidos, la revolución digital en sus diferentes expresiones ha ido avanzando velozmente sin que la sociedad occidental haya tomado debida nota, la cual de repente, en la tercera década del siglo XXI, se encuentra ante un escenario en la que de ser el director de la obra pasó a ser un débil actor de reparto, donde no sabe como se llama la obra, no sabe cuál es el tema de la obra y desde luego tampoco adivina cuál será su final y desenlace.

La increíble rapidez de lo digital, donde ahora los chips de inteligencia artificial se mejoran y capacitan cada 24 hrs., crea un escenario de locura donde este Estado débil no sabe exactamente lo que está pasando.

Al propio tiempo se ha difundido por los dueños del poder digital que este movimiento es *inevitable* y todo aquel que se oponga es un ludita anacrónico que se está oponiendo a ese deslumbrante futuro, que sin decirnos cuál es, el poder digital difunde todos los días en todos lados.

Se argumenta bajo criterios nuevamente neoliberales, que a una menor regulación le corresponde un mayor poder digital, así como el crecimiento imparable de una innovación tecnológica que no debe ser tocada por ningún

38 Zuboff, Shoshana (2020), Op. Cit., p. 31.

Estado porque de hacerlo estaría deteniendo el futuro de la inteligencia artificial con el consiguiente regazo en la competencia geopolítica mundial.

Al propio tiempo, el Estado moderno carece del conocimiento e instrumental necesarios para entender y administrar toda la innovación que fluye desbordada todos los días sin ningún recato, lo que ha llevado a la paradoja de no poder regular lo de hoy porque no tengo regulado lo del año pasado, lo cual me limita a regular lo que tendré la semana que viene; a lo cual se suma una producción imparable de novedades digitales en chips, nanotecnologías, tecnología 5G, bandas anchas, etc., y versiones distintas de inteligencia artificial que salen al mercado sin haber sido evaluadas por ninguna autoridad y de las cuales no se saben sus implicaciones o sus efectos respecto a los individuos o al agregado social, creando una catarata de datos que no solo ahoga cualquier intento de regulación, si no que también impacta al equilibrio y estabilidad de los grupos sociales[39].

Aprovecha también que el debate sigue instalado en un punto de datos privados o no; si deben ser pagados o no, y que en ese reduccionismo se ha renunciado a generar un momento de reflexión que yendo al centro de lo que representa esta nueva revolución de lo digital, pudiera pensarse por filósofos, historiadores, políticos, economistas, por abogados, etc., que no sólo resolvieran si el dato pesonal puede ser violado impunemente; si no que también propiciará el inicio de un nuevo relato de amplias dimensiones culturales en el que el ser humano pudiera prever con cierto grado de racionalidad que es lo que

39 Al respecto Sadin señala que "El segundo factor que favorece esta extensión incesante se origina en el hecho de que la producción industrial actual no respeta una serie de fases que hasta hace poco tiempo estaban marcadas por diversas formas de indeterminación dentro de las búsquedas, o por la aceptación del fracaso como riesgo consustancial a la elaboración de cualquier prototipo, o incluso por la exigencia de tener que proceder a múltiples y minuciosos testeos de calidad- que vemos que cada vez más hacen falta-. Hoy, el ritmo marca la casi ausencia de lapso temporal entre la concepción y la comercialización de los productos mismos. La presión de la competencia y la primacía del retorno inmediato después de cualquier inversión prohíben toda latencia, toda evaluación concertada sobre el valor y la pertinencia de los productos. Las unidades de investigación y desarrollo deben probar sin demora y sin pausa que representan el acceso a nuevos beneficios. Estamos. En el momento de los "ciclos de innovación", que siempre están más cerca unos de otros en la medida en que favorezcan una dinámica de entusiasmo permanente y que estén investidos por los dogmas conjugados del crecimiento y del aumento del supuesto confort de las personas (que, por el bien de la sociedad, no deben nunca dejar de crecer).(Sadin, Éric. (2023) *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical*, Caja negra, pp. 22-23).

está pasando, cuales son los riesgos y las ventajas de la inteligencia artificial y cuales son las decisiones que debe tomar no sólo una persona o un país, si no una humanidad que está emplazada a traducir sus grandes avances en ciencia y tecnología en una victoria o beneficios, sin perder en ello la libertad o su sentido de la vida.

Lo que se aprecia es que ante la crisis del debate liberal, una sociedad global que no acaba de entender su nuevo papel en la historia, presenta una amplia desventaja epistemológica para intentar el inicio de un nuevo relato que esté a la altura del nivel científico y tecnológico al que ha llegado a través de diez mil años de historia.

Debe decirse por otro lado que hasta el día de hoy los diversos intentos por regular esta nueva ola digital han sido limitados. Como se sabe, Europa ha registrado el mayor avance en la materia, al igual que un intenso debate en el que se le presiona señalando que su nueva regulación digital es la razón de su atraso tecnológico a manera de advertencia para otras regulaciones del mundo. Por ejemplo, la Regulación de Protección de Datos Personales de la Unión Europea (GDPR por sus siglas en inglés) de 2018, no ha sido lo exitosa que se esperaba en razón de no existir una idea clara sobre lo que se protege y de precisar hasta donde un dato personal puede ser utilizado en contra de uno mismo; o lo que pasa constantemente, de que las diferentes redes de datos brindan una información personal indirecta proveniente de la vida social de cada uno de los usuarios, lo cual ha derivado que esta regulación afecte de manera más fuerte a las empresas pequeñas y solo tangencialmente a los grandes monopolios como Google, Facebook y Microsoft, etc.

En el caso de la inteligencia artificial, en 2024 la misma Unión Europea aprobó una ley en la materia (Ley de IA), la cual desde su salida generó una fuerte polémica por estimarse nuevamente que su regulación limitaba la operación de sus innovaciones; y por el otro, los críticos del capitalismo de la vigilancia estimaron que su contenido era débil y sus alcances limitados. Como lo mencionó Max Von Thun, director para Europa del Open Markets Institute "hay importantes lagunas jurídicas en la ley para las autoridades públicas" y "una regulación relativamente débil de los modelos de base más grandes que plantean los mayores daños"[40].

No obstante lo anterior, la plataforma regulatoria de la Unión Europea integrada hoy por más de cuatro leyes en la materia, se erige como una referencia obligada para las demás naciones del mundo.

40 Euronews, 18 marzo 2024.

En lo que respecta a Estados Unidos, cuna de la política neoliberal y país sede de las mayores empresas digitales del mundo, desde 1995 estas empresas han venido poniendo a salvo sus intereses al promover en ese año el cierre de la Oficina Norteamericana de Activos Tecnológicos (OTA por sus siglas en inglés), la cual se había instituido desde 1972 a través de un grupo multidisciplinario que intentaba darle transparencia y control a los avances tecnológicos. De igual modo en 1998, a través de la sección 230 del Acta de Comunicaciones y Decencia, se estableció una amplia protección para las plataformas de internet en contra de cualquier intento de regulación sobre el contenido almacenado en dichas plataformas, donde se estipuló que "ningún proveedor o usuario de servicios computacionales interactivos será tratado como el editor o comunicador de cualquier información proporcionada por otro proveedor del contenido de la información"[41].

Este criterio de libre mercado en favor de las grandes compañías de tecnología en Estados Unidos fue ratificado recientemente por el Tribunal Supremo de Estados Unidos el 01 de julio de 2024, cuando determinó a través de sentencia que los propietarios de plataformas sociales no podían ser regulados de manera externa, por lo que solo procedía la autorregulación de sus contenidos[42].

A partir de la llegada de Trump al poder esta tendencia se ha profundizado al emitir las siguientes órdenes ejecutivas:

- Orden ejecutiva sobre la libertad de expresión

  Trump emitió una orden ejecutiva titulada **«Restaurando la libertad de expresión y poniendo fin a la censura federal»**, que afirma que la Administración Biden "pisoteó los derechos de libertad de expresión al censurar el discurso de los estadounidenses en plataformas en línea", bajo el pretexto de combatir la desinformación y la información errónea. La orden instruye al Fiscal General a investigar las políticas federales de los últimos cuatro años y recomendar acciones correctivas adecuadas.
- Memorando para limitar nuevas regulaciones

  En un memorando titulado «**Congelación regulatoria pendiente de revisión**», Ordena a todas las agencias ejecutivas que no propongan nuevas normas hasta ser revisadas y aprobadas por un jefe de departamento, lo que podría frenar la regulación digital y de IA.

41 Acemoglu, Daron, (2023) Op. Cit. p. 412.

42 El país, 01 Julio 2024.

- Orden ejecutiva sobre la postergación de la prohibición de TikTok
  Titulada «Aplicación de la Ley de Protección de los Estadounidenses de Aplicaciones Controladas por Adversarios Extranjeros a TikTok», efectivamente retrasó la prohibición en EE.UU. de la aplicación china de videos cortos TikTok por 75 días.

De igual modo, vale la pena señalar que a partir de la llegada de Trump al poder, Musk, Zuckerberg y otros grandes consorcios digitales abandonaron el uso de verificadores de datos independientes en plataformas como Facebook, Instagram, X, etc., reemplazándolos por el sistema de "notas de comunidad". Esta decisión ha generado preocupación sobre el aumento de la desinformación, ya que en lugar de contar con profesionales que verifiquen la veracidad de los contenidos, ahora dependen de la evaluación de los propios usuarios; un mecanismo que puede ser fácilmente manipulado por campañas organizadas, intereses políticos o cuentas falsas[43].

VIII. REFLEXIONES GENERALES

Resulta evidente que somos pasajeros de una etapa estelar de la historia de la humanidad.

Bajo el enfoque que se acepte, con la profundidad que se vea; una primera coincidencia podría ser que la llegada de la revolución digital nos invita como humanidad a repensarnos como sociedad humana en el marco de nuestra historia, con todas nuestras fortalezas pero al mismo tiempo con nuestras debilidades.

En este sentido, exagerar sobre la conciencia del momento que se vive no sería más que intentar ubicarse lo más cercano posible a un fenómeno de dimensiones estelares que ya está cambiando de manera disruptiva e ininteligible la vida personal, familiar, social y política de la humanidad.

La aparición de los discursos de alerta que empiezan a escribirse en este siglo, resultan por demás valiosos ante la ausencia de una bibliografía que explique de manera suficiente la aparición de un fenómeno digital que carece de relato.

A lo largo de este pequeño ensayo, lo que hemos intentado subrayar es que junto con la llegada de una nueva era digital aparece la debilidad de un relato

43 El País, 12 de Enero 2025.

liberal que hasta el siglo pasado había sido suficiente para que la sociedad occidental debatiera las distintas formas de su identidad y progreso.

Que el quebranto de un modelo neoliberal, bajo la idea de la desaparición o adelgazamiento del Estado por cerca de medio siglo, ha dejado por herencia una insatisfacción social y un vacío de poder que ha resultado en el mayor beneficio de los dueños de la era digital.

Al ir quedando desprovistos de una plataforma liberal que nos daba rumbo y sentido, están apareciendo en el campo geopolítico, económico y social, diversas voces que intentan llamar a la plaza pública a debatir el tema.

Desde luego, junto con estas voces de alarma aparece una corriente muy poderosa encabezada por los propios monopolios de la tecnología, que amparados por las grandes luces que emanan de la inteligencia artificial, encabezan el desfile de un gran ejército de la población mundial que sin bandera, sin relato y sin objetivos claros, conducen a la humanidad a una *utopía distópica* que no es posible hoy dibujar. Autores como Kai-Fu Lee, Bostrom, Floridi, Tegmark, Schmidt, etc., nos entusiasman con sus arengas sobre una inteligencia artificial que a la fecha ya está dando grandes frutos.

El brillo que emana de estos logros del intelecto humano, concretados por las primeras etapas de la inteligencia artificial, se convierten en alimento y muchas veces en adicción de una parte de la sociedad que sin reparos; seducida ante las innovaciones digitales e ilusionada por ese futuro de inmortalidad, de vejez eterna, de perfección física y de felicidad; va cayendo en una enfermedad datológica que la va trasladando a una libertad asistida donde paulatinamente va perdiendo su capacidad de decidir.

Por otro lado, aparecen autores como Byung-Chul Han, Lassalle, Zuboff, Suleyman, Acemoglu, Johnson, Susskind, Noah, y otros que tratando de alejarse de la fascinación digital, se han colocado en una posición crítica respecto a los dueños del avance tecnológico, denunciando que se ha perdido mucho tiempo en tomar las decisiones correctas de una regulación digital en favor de la humanidad.

Autores como Lassalle nos hablan de un tecnopoder que socava ya los fundamentos de equidad de la democracia liberal, al promover una estructura algorítmica de la sociedad que influye en el ejercicio de la libertad de los seres humanos sin debate público ni legalidad. De igual modo nos advierte de la llegada de un ciberleviatán en marcha que se allanará sin violencia, si no se hace algo a tiempo. Agrega que cuando la ciudadanía admita que ha dejado de ser el sujeto de la soberanía de forma natural y sin resistencias, caerá en un ciberpopulismo de redes sociales. En ese nuevo orden de lo digital, el pueblo

se disolverá en un proletariado cognitivo que adoptará el rol de consumidor masivo de conectividades tecnológicas[44].

Estos conceptos por demás audaces, para un público escéptico, se fundamentan entre otras razones en el ejercicio de un neoliberalismo que siempre soñó con la desaparición del Estado, y que su posible extinción ahora a través de un nuevo poder digital, no le parece reprobable; aludiendo incluso al caso de tecnólogos hipelibertarios como el fundador de Paypal y capitalista de riesgo Peter Thiel, Mounk, y otros.

Mustafa Suleyman, cofundador de Deepmind, una de las empresas líderes en inteligencia artificial a escala internacional, no obstante de ser uno de los grandes impulsores de este movimiento, de manera reciente ha referido que la tecnología surgida de la inteligencia artificial no es un invento más dentro del amplio abanico de logros científicos que el ser humano ha conseguido a lo largo de su historia. Por ello señala que "Necesitamos con urgencia respuestas indiscutibles sobre como contener y controlar la ola que viene, sobre como sería posible mantener las salvaguardas y las posibilidades del estado nación democrático – aunque reconoce – que por ahora, nadie tiene un plan así"[45].

Acemoglu comenta que desafortunadamente la democracia no es compatible con los nuevos escenarios de la ilusión de la inteligencia artificial manejados por las grandes compañías tecnológicas; que al contrario, que son opuestos en cuanto al manejo de la libertad y la voluntad algorítmica de la inteligencia artificial y de sus intereses, lo cual debilita la democracia cuando más necesidad se tiene de ella[46].

Susskind contribuye al perfil de este nuevo poder digital, aduciendo que no es moralmente neutro u objetivo; que tiene una naturaleza política y que dada su novedad, no reside en palacios de gobierno o parlamentos; que opera fuera de los canales de la política tradicional profanando la libertad y deteriorando la democracia bajo un grupo de fuerzas que no pueden ser atribuidas a una sola corporación o persona[47].

Zuboff agrega que el capitalismo de la vigilancia no es una tecnología, que es una lógica que impregna la tecnología y que la pone en acción. Y deja claro, haciendo patente un sentir de la mayoría de estos autores, que es el capitalismo de la vigilancia y no la tecnología lo que pone precio a la subyugación y a la impotencia, al moldear el comportamiento para que favorezca los resultados

44 Lassalle, José María (2019), Op. Cit. pp. 95-120.
45 Suleyman, Mustafa (2024), Op. Cit. p.26.
46 Acemoglu, Daron y Johnson, Simon (2023), Op. Cit. pp.372-380.
47 Susskind, Jamie (2022), *The Digital Republic*, Pegasus, pp.4-5.

comerciales (y ahora políticos) de los capitalistas de la vigilancia, debilitando con ello a la política y a la democracia[48].

Noah declara que "en su forma actual, la democracia no sobrevivirá a la fusión de la biotecnología y la infotecnología. O bien se reinventa a si misma con éxito y de una forma radicalmente nueva, o bien los humanos acabarán viviendo en dictaduras digitales"[49].

• DE CONTENCIÓN Y OTROS REMEDIOS.

El avance del poder digital ya es notorio en el campo económico ante el tamaño que han adquirido las grandes empresas de tecnología. En materia política, ante su intervención cada vez más irreverente en la venta del mercado de datos, como la propia intervención de los CEO's en los diversos procesos políticos no solo de Estados Unidos, sino de otras partes del mundo. Y en la parte social, ante la acumulación de los datos de miles de millones de seres humanos que al día de hoy si estuvieran en CD's podrían darle la vuelta al mundo 222 veces.

A esto habría que agregar el proceso que ya se ha echado a andar con base a estos datos, de un esquema de aprendizaje (machine learning) de una inteligencia artificial que trabaja las 24 horas del día los 7 días de la semana en su mejora generativa y en consecuencia de todos los procesos que se le han encargado en los diferentes campos de la humanidad, ya sean de salud, del área militar, urbanos, económicos, sociales, infraestructuras, biología sintética, espacial, etc.

La disyuntiva del tema no es fácil. En este momento las opiniones se dividen entre los que señalan que la inercia actual está dominada por un mundo sin regulación digital que favorece a las 5 o 10 empresas que en Estados Unidos y el mundo controlan este poder digital. También se habla de que este desarrollo digital ya está inmerso en una guerra geopolítica sobre el avance y control de la inteligencia artificial que como lo señalara el presidente Ruso Vladimir Putin "el que controle la inteligencia artificial controlará el mundo".

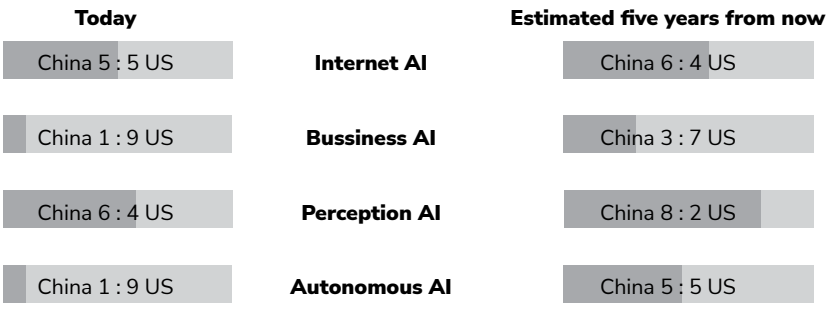
La consideración de esta opinión en el marco de dos guerras que a nivel mundial se suceden hoy en Ucrania y Palestina, lo vuelve un argumento creíble; el cual se vuelve más fuerte si agregamos el movimiento de retoma del poder euroasíatico; así como el choque que Estados Unidos y China vienen sosteniendo de manera abierta desde 2016 por el liderazgo mundial.

48 Zuboff, Shoshana (2020), Op Cit., p. 30.

49 Noah, Yuval (2018), Op. Cit. p.89.

Siendo los dos grandes líderes de la inteligencia artificial, toda consideración que vaya en el sentido de administrar este avance del nuevo poder digital, tendría que pasar por el acuerdo geopolítico de las dos naciones que hoy detentan el 90% del poder de la inteligencia artificial del mundo.

GRÁFICA 4 Carrera por el liderazgo en la IA



The balance of capabilities between the united States and China across the four waves of AI, currently and estimated for five years in the future

FUENTE: AI Superpowers, Kai Fu – Lee p. 136

Junto con ello, como se indicó en párrafos anteriores, dada la magnitud de lo que el tema representa para el presente y el futuro de la humanidad, sería deseable la celebración de una convocatoria lo más amplia posible, a fin de que bajo las diferentes ópticas con las que hoy se explica al ser humano, entiéndase la filosófica, histórica, jurídica, psicológica, económica, social, política, ecológica, entre otras, se reflexionara con rigor y seriedad, más allá de la fascinación y del espanto, sobre como utilizar este gran avance de la inteligencia artificial para beneficio de la mayor parte de la humanidad[50].

50 El advenimiento de la IA, con su capacidad para aprender y procesar información de maneras que la razón humana por sí sola no logra, puede producir avances en cuestiones que han demostrado estar más allá de nuestra capacidad para responder; pero el éxito generará nuevas preguntas. La inteligencia humana y la inteligencia artificial se están encontrando y aplicando a actividades a nivel nacional, continental e incluso global. Comprender esta transición y desarrollar una ética orientadora para ella requerirá el compromiso y la percepción de muchos elementos de la sociedad: científicos y estrategas, estadistas y filósofos, clérigos y ejecutivos, CEO's y empresarios. Este compromiso debe hacerse tanto dentro de las naciones como entre ellas. Ahora es el momento de definir tanto nuestra asociación con la inteligencia artificial como la realidad que resultará de ella (Kissinger, Henry, 2021, *The The age of A.I. And our human future*, USA, p. 226.)

De igual modo, sería deseable llegar a una serie de recomendaciones generales que pudieran motivar el compromiso de las naciones para administrar, reglamentar y encauzar a la inteligencia artificial.

En este punto surge la inquietud de algunos científicos o especialistas, sabiendo de antemano la enorme dificultad que esto representa. Sin embargo, como lo sintetiza Suleyman "A simple vista, la contención (de la inteligencia artificial) no es posible, pero por el bien de todos, es imprescindible que lo sea"[51].

En esta negativa a priori de algo que una mayoría no ve posible de lograr por sus implicaciones geopolíticas y económicas, estriba el reto de una sociedad de la tercera década del siglo XXI, para que en un momento de claridad pudiera darse una oportunidad de trascender hacia un futuro de lucidez en el marco de la Primera Revolución Digital.

Aparecen al alcance de la sociedad global un sinnúmero de alternativas que podrían contribuir a la construcción de un liberalismo digital. La implementación de nuevos acuerdos globales, regionales o nacionales, que en el marco de sus circunstancias puedan aplicar medidas de remedio.

Entre estas opciones destacan alternativas como atender a incentivos de mercado, inhibir políticas de vigilancia; ralentizar el desempleo tecnológico; pagar por el uso de datos consentidos, control a las grandes Big Techs; ralentizar la automatización; invertir en los trabajadores; recuperación del poder del Estado; protección de datos; legislar sobre responsabilidad de contenidos; impuestos a los anuncios digitales; educación digital; compartimiento de tecnologías; tribunales especiales y otros sinnúmero de medidas que en vez de que nos lleven a un ciberleviatán, nos puedan conducir a una República Digital donde de alguna manera se preserve la libertad, la democracia y el Estado de Derecho, bajo nuevos conceptos que permitan una civilización digital del siglo XXI.

El tema de la nueva era digital es un relato abierto donde se empieza a discutir, con conciencia o no, el futuro de la humanidad y de la especie humana como responsable de su destino.

Hay antecedentes, pero ninguno de sus escenarios nos confrontó como el de ahora.

En los próximos días, en los próximos años, tendremos que decidir obligadamente entre la estelaridad o la intrascendencia.

51 Suleyman, Mustafa (2024), Op. Cit., p.28.

IX. BIBLIOGRAFÍA

Acemoglu, Daron y Johnson, Simon (2023), Power and Progress, Publicaffairs. USA.

Carr, Nicholas (2014), Atrapados Cómo las máquinas se apoderan de nuestras vidas, Taurus.

Crawford, Kate (2021), Atlas of AI, Yale. USA

El país, 01 Julio 2024.

El País, 07 de Agosto, 2024.

El País, 12 de Enero 2025.

Euronews, 18 marzo 2024.

Fondo Monetario Internacional (PIB- Per Cápita Mundial (0-2000)

Han, Byung-Chul (2022), Infocracia, Penguin Random House, México.

Innenarity, Daniel (2020), Una teoría de la democracia compleja Gobernar el siglo XXI, Gutenberg. Barcelona.

Kissinger, Henry, 2021, *The The age of A.I. And our human future*, USA.

Lassalle, José María (2019), Ciberleviatán, Arpa. Barcelona.

Lassalle, José María (2021), El liberalismo herido, Arpa. Barcelona

Lee, Kai-Fu (2018), AI Superpowers: China, Silicon Valley, and the New World Order, Harper Business, USA.

López Portillo, José Ramón; Blog Nexos, julio 2024.

Luce, Edward (2017), The retreat of Western Liberalism, Atlantic Monthly Press. USA.

Market Cap. El país, 27 de julio 2024.

Mayer, Colin (2024), Capitalismo y Crisis. Como arreglarlas, Oxford. Reíno Unido.

Mounk, Yascha (2018), El pueblo contra la democracia, Espasa. España

Noah, Yuval (2018), 21 lecciones para el siglo XXI, Debate. México.

Oropeza, Arturo (2024), Hacia una nueva visión del desarrollo industrial digital en México, México.

Research Gate - Evolución de la población mundial (1500- 2050)

Rifkin, Jeremy (1996), El fin del trabajo, Grupo Planeta.

Rose, Matthew (2021), A world after liberalism, Yale. USA.

Sadin, Éric. (2023) La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical, Caja negra. Buenos Aires.

State Counter, El país, 27 de julio 2024.

Stiglitz, Joseph (2024), The road to freedom, Norton Company, USA.

Suleyman, Mustafa (2024), La ola que viene. Debate. México

Susskind, Jamie (2022), The Digital Republic, Pegasus. USA.

Vanguadia, Dossier (2020).
Vanguardia 2020
Wolf, Martín (2024), La crisis del capitalismo democrático, Arial. México
Zuboff, Shoshana (2020), La era del capitalismo de la vigilancia, Paidós. México.

INTELIGENCIA ARTIFICIAL: ENTRE LA DISRUPCIÓN Y LA REGULACIÓN

SUMARIO:

I. Tecnología digital: la era de los paradigmas **II.** Renacimiento, chips e inteligencia artificial **III.** Entre el miedo y la fascinación digital: el síndrome de Diego Rivera **IV.** La era Tump y el corporativismo digital **V.** Conclusiones. **VI.** Bibliografía

> JORGE FERNANDO NEGRETE
Presidente de Digital Policy & Group. Ex presidente de la Asociación Mexicana de Derecho a la Información A.C. (AMEDI).

I. TECNOLOGÍA DIGITAL: LA ERA DE LOS PARADIGMAS

En materia de política de ciencia y tecnología, hablamos de una ola tecnológica cuando una tecnología en específico tiene la capacidad de masificar el uso de cientos o miles de otras más. El vapor, la energía eléctrica, la industrialización, como un fenómeno de tecnologías complejas, son un ejemplo de esto. Pero desde hace 50 años, la tecnología digital y la conectividad han tenido la capacidad de impactar no sólo como una ola, sino en el territorio de los paradigmas de nuestra sociedad, de forma multidimensional, construyendo un fenómeno que, en el territorio de la epistemología, nos lleva a un pensamiento complejo.

Existe un nuevo paradigma en la sociedad cuando un hecho tiene la capacidad de transformar, modificar e impactar el marco legal, económico, social y cultural de una sociedad. Desde esa perspectiva, la tecnología digital ha sido disruptiva, con un impacto visceral, y ha tenido la capacidad de crear estos paradigmas que transforman compulsivamente nuestra sociedad. El Internet, la banda ancha, la convergencia, la nube, el cómputo, el Internet de las cosas, la realidad virtual y aumentada y, de forma reciente, la inteligencia artificial (IA), son las tecnologías, hijas de la digitalización, que hoy están presentes y cada una de ellas es por sí sola, un nuevo paradigma en nuestra sociedad digital.

LA OLA DE LA CONECTIVIDAD.

Pero, ¿cuál ha sido el elemento masificador de estas tecnologías en el mundo? Las telecomunicaciones o, lo que hoy se denomina infraestructura digital. Todos estos paradigmas sin las redes de telecomunicaciones existirían, pero de forma acotada, sin la eficiencia de su portabilidad e incorporación en la vida de las personas.

Las redes de telecomunicaciones masificaron estas tecnologías, nos permitieron aproximarnos a ellas y habilitaron millones de soluciones digitales. Personalidades como el Ing. Carlos Slim Helú han señalado que las telecomunicaciones son "el sistema nervioso de nuestra civilización", y tiene razón. La infraestructura digital, particularmente la generación denominada 4G, transformó lo que hoy es nuestra sociedad digital. Surgieron cambios profundos y externalidades inimaginables:

1. Las redes de conectividad se volvieron mercados.
2. Estos mercados se volvieron globales.
3. Los hábitos de consumo se transformaron. El consumo en puntos de venta territorial específico se transformó en un consumo de bienes y servicios, deslocalizado y concentrado en el equipo de transmisión: computadora, *smartphone*, tableta, Smart TV, etc.
4. Ocurrió la desmaterialización de ciertos productos físicos o editoriales.
5. La era del acceso. Todo mundo quiere el acceso a fondos de catálogos de derechos de autor y formatos de consumo como las plataformas de video, música y textos. Nacen empresas como Netflix, Amazon Prime Video, Spotify, Disney+, Tidal, Kindle, etc.
6. Nace la economía colaborativa. Muchas personas ya no quieren un auto, sino el acceso al transporte vía una aplicación como Uber o DiDi. ¿Tienes un auto, patines, moto o bicicleta?, entonces puedes incorporarte a una plataforma que intermedia la venta de productos, servicios y su logística y entrega.
7. Surgen nuevos modelos de negocio y cadenas de valor en línea.
8. La innovación se transforma en una conducta inevitable y permanente.

ESTADO, PERSONA DIGITAL Y CONSTITUCIÓN

El derecho de acceso a Internet habilita el ejercicio de los derechos fundamentales en el mundo digital. Es un derecho de apertura, poderoso habilitador de una realidad virtual donde se vive plenamente una vida de derechos y economía digital. El ciudadano digital, siempre conectado, en todas partes, adquiere mayor relevancia al encarnar las diversas tecnologías digitales y usarlas en

su vida cotidiana. El cómputo, los procesadores, la industria del *software*, la televisión de paga, los contenidos de propiedad intelectual en sus formatos musicales, audiovisuales y libros, no existirían sin la infraestructura digital. Las telecomunicaciones están en la primera línea para abatir la exclusión digital, es la puerta a un universo de derechos humanos y economía digital.

¿Existe una relación entre Estado, Constitución, derechos fundamentales, tecnología digital y política pública? Sí. ¿Son un derecho fundamental el acceso a Internet, a la tecnología digital y los servicios de telecomunicaciones? Sí. ¿El Estado y sus instituciones deben proteger, promover el acceso a Internet, las tecnologías digitales y la infraestructura digital? Sí. ¿Diseñamos la regulación y política pública que necesitan nuestros ciudadanos? No siempre. ¿La política pública y la regulación habilitan el ejercicio de todos los derechos fundamentales? Si se hace bien, sí. Si no, conculca derechos, a veces por años.

El Estado y los poderes públicos son creados por los seres humanos para poder convivir y son, esencialmente, una entidad jurídica. El Estado no sólo es el territorio dotado de un poder de mando originario (Heller), ni el espacio de validez de las normas jurídicas (Kelsen). El Estado es un complejo creado y estructurado por el orden jurídico, territorio, derechos fundamentales, gobierno y acción pública.

En su concepto estático, el Estado se estructura en la Constitución; y en su concepto dinámico, es la base de la acción pública, sus órganos, funciones y regula la vida social. Es aquí donde hablamos de derechos fundamentales, en un caso, y de gobierno y política pública en otro.

Ahora bien, el constitucionalismo es un enfoque jurídico y político sustentado en la irrestricta sujeción del gobierno y las instituciones del Estado a la Constitución y sus mandatos. En este caso, un mandato novedoso: garantizar el acceso a Internet, las TIC y los servicios de telecomunicaciones para ejercer los derechos fundamentales en el mundo de la tecnología digital.

Sostengo que el derecho de acceso a Internet es un derecho humano que posee toda persona para conectarse a Internet por medio de las tecnologías digitales y la infraestructura digital, con el fin de ejercer y disfrutar de sus derechos fundamentales. El Estado tiene la responsabilidad de garantizar el acceso a este servicio y tecnologías. Estar siempre conectado en todas partes y gestionando el ejercicio pleno de nuestros derechos fundamentales y economía digital, permite el nacimiento de la persona digital.

El Consejo de Derechos Humanos de la ONU reconoció en 2012 la protección y disfrute de los derechos humanos en Internet. Este documento fue suscrito por un grupo relevante de países en Iberoamérica: Chile, España y México, entre otros.

La visibilidad que se da a la Internet, en términos del lenguaje de derechos humanos, y el ejercicio de otros derechos fundamentales bajo esta óptica por medio de la tecnología, ofrece una nueva dimensión de análisis en términos constitucionales. Hablamos de derechos universales, interdependientes, indivisibles y programáticos: salud, educación, cultura, libertad de expresión, entre otros. No son derechos superiores uno sobre otros, interactúan directamente entre ellos y su cumplimiento se verifica en el tiempo. Por eso, quien está conectado a Internet y a un dispositivo móvil o fijo preserva el ejercicio de sus derechos fundamentales: libertad de expresión, derecho a la información, acceso a la cultura, educación, movilidad y economía digital.

Ninguna autoridad puede restringir directa o indirectamente el acceso de una persona a Internet, TIC y servicios de telecomunicaciones. La restricción es una violación flagrante al ejercicio de todos los derechos fundamentales. Nace el "constitucionalismo digital", incluso un constitucionalismo digital con vocación iberoamericana: México, España y ahora Chile debaten una nueva generación de derechos, los digitales.

El "constitucionalismo digital" escribe sus primeras páginas y se nutre de una agenda transversal de oportunidades, donde cada derecho fundamental es el centro de nueva regulación y política digital, bajo la primera revolución digital de nuestra civilización.

II. RENACIMIENTO, CHIPS E INTELIGENCIA ARTIFICIAL

La frontera digital no está rumbo al cielo, sino hacia lo diminuto. No hay Inteligencia Artificial (IA) sin procesadores y en esta tecnología se presenta una de las grandes batallas tecnológicas de nuestra historia. No existe IA ni poder político, ni económico, ni sociedad digital sin procesadores de cómputo (*chips*).

Los semiconductores, procesadores o *chips* de cómputo son la frontera de todo, la semilla primigenia de la sociedad digital. Sin ellos no hay cómputo, procesamiento de datos, IA, servicios digitales ni transformación digital. Son el principio y la frontera tecnológica de la sociedad digital. Estados Unidos, en su corta historia, ha dado las batallas tecnológicas más importantes de nuestra sociedad contemporánea. ¿La más reciente? La batalla de los *chips*, los procesadores en un entorno que habilita el nacimiento de la IA.

El poder de esta industria bautizó una región geográfica y económica de EUA: Silicon Valley, y nació bajo el éxito, primero militar, después de la guerra de Vietnam y más tarde con el mercado de bienes electrónicos de consumo.

Ha enfrentado dos veces a Asia. Primero frente a los japoneses a finales de los ochenta y mediados de los noventa, con la poderosa industria electrónica de Fuji, Canon y Sony, y ahora frente a China. En ambas ocasiones, una ley de procesadores desde el Congreso le regresó a EUA su liderazgo tecnológico digital.

Todo producto digital tiene un procesador o *chip* como corazón. Bienes electrónicos, de consumo, autos, aviones, computadoras, barcos, armas militares, radares, *smartphones* y la industria aeroespacial. De los *chips* depende la seguridad de Estados Unidos, Occidente y el mundo.

Los procesadores son la base del procesamiento de datos, del cómputo y el supercómputo. Las redes de telecomunicaciones, la infraestructura digital, tienen un solo propósito, facilitar el transporte de masivas cantidades de datos y procesos de cómputo. Nunca tan juntos *chips* y redes.

La Inteligencia Artificial es hija de padre y madre: procesadores y redes de telecomunicaciones o infraestructura digital. No hay Inteligencia Artificial ni redes robustas de telecomunicaciones sin *chips*.

Durante décadas, los transistores, empalmados a tarjetas electrónicas grandes, se concentraban por cientos y miles dentro de pisos y cuartos de cómputo. Hoy, la nueva generación de *chips* puede colocar más de 290 millones de transistores por milímetro cuadrado. Empresas como Intel pretenden fabricar *chips* de un millón de millones de transistores para 2030. Hablamos de procesadores de 3NM (nano milímetros). La ciencia ficción vía la tecnología de frontera llamada litografía ultravioleta profunda, LUP.

Las empresas de procesadores pretenden que la IA se considere *everywhere*, ubicua, que llegue a la Nube, la computadora portátil, al cómputo al borde o Edge Computing (o como le llamo, Data Center de barrio) y dispositivos móviles. Esto ya desata la mayor y vertiginosa carrera por el control de la Inteligencia Artificial vía el diseño de procesadores de cómputo, denominados neuromórficos, semejando la capacidad neuronal humana.

El ex presidente Joe Biden adicionó a la ecuación la identificación de masivas cantidades de espectro radioeléctrico a la ecuación. Las empresas de procesadores no dudan en llamar a este momento como una "economía del silicio" que vale 600,000 millones de dólares y una probable aportación de un trillón de dólares a la economía global en 2030.

El axioma es: redes de telecomunicaciones + chips neuromórficos + espectro = capacidad de cómputo = Inteligencia Artificial.

Estados Unidos ha limitado esta tecnología a China. El planeta entero busca la oportunidad de generar las condiciones para ser atractivos y llevar inversiones

estratégicas a sus regiones. Taiwán ya no es un puerto seguro en inversión y maquila de procesadores.

Al centro de esta frontera se encuentran las acciones de los gobiernos. Hablamos de dinero público, Estado y rectoría económica. Aquí, algunas acciones concretas para promover inversiones en materia de procesadores y diversificar su producción.

Estados Unidos. La Ley de Ciencia y CHIPS (Creating Helpful Incentives to Produce Semiconductors), promovida por el presidente Joe Biden, establece un proyecto estratégico de seguridad nacional y pone a disposición 280 mil millones de dólares para reforzar la competitividad en el sector tecnológico, hacer frente a la creciente amenaza tecnológica de China y proteger al país de posibles disrupciones a la cadena de valor. De esta cantidad, dirige 52 mil millones de dólares para la investigación y desarrollo de procesadores de cómputo.

Europa. El Consejo Europeo aprobó la regulación para reforzar el ecosistema regional de semiconductores, conocido como Chips Act, mediante la cual busca generar una inversión de hasta 43 mil millones de euros para impulsar la fundición de 20% de *chips* del mundo, en territorio europeo.

España. Prevé movilizar una inversión pública en procesadores de hasta 12,250 millones de euros.

Alemania. Prepara un fondo por 20 mil millones de euros para impulsar la manufactura de *chips* en el país, mediante la entrega de subsidios a empresas hasta 2027.

México. Ya es un socio estratégico en la cadena de valor de la industria de los procesadores. La empresa Intel tiene su centro de diseño global de procesadores en Guadalajara y, recientemente, la empresa Foxconn anunció que pondrá una fábrica para fabricar los procesadores de IA de Nvidia.

Costa Rica. El Secretario de Estado, Anthony Blinken, anunció una alianza estratégica con el gobierno de Costa Rica para explorar oportunidades y hacer crecer el ecosistema global de semiconductores y el nuevo Secretario de Estado de EUA, Marco Rubio, ratificó el apoyo a Costa Rica como un país aliado y socio estratégico en la región.

Panamá. Estados Unidos y Panamá colaborarán para impulsar el sector de los semiconductores y evaluar el ecosistema de semiconductores existente en el país centroamericano, a fin de identificar las fortalezas y áreas de mejora del país dentro del sector.

Por otra parte, 2024 será recordado como el año de los procesadores y la Inteligencia Artificial (IA).

- TSMC (Taiwan Semiconductor) anunció una inversión de 40 mil millones de dólares en Arizona, y Samsung anunció que construirá en Taylor, Texas, una planta de *chips* avanzados, en la cual invertirá 17 mil mdd.
- Nvidia se consolidó como el líder en el campo de la tecnología que hace factible la IA y el cómputo extremo. Nvidia tuvo un aumento en el valor de sus acciones, superando por momentos a Google, Apple, Microsoft y convirtiéndose en la empresa más valiosa del mundo en la Bolsa.
- TSMC, Taiwan Semiconductor Manufacturing Company, tuvo un 2024 único en su historia, consolidándose como el mayor fabricante de *chips* del mundo. Lideró la tecnología de 3 nanómetros (nm), la cual representa la vanguardia en la fabricación de semiconductores, más pequeños, potentes y eficientes en energía en el mercado de *chips* para aplicaciones como IA y dispositivos móviles de última generación.
- AMD tuvo un 2024 muy exitoso, consolidándose como un fuerte competidor en el mercado de CPUs, GPUs y logró un récord de 33.9% del mercado de servidores.
- Mediatek desarrolló *chips* de 3 nm y lideró Wi-Fi 7, lo cual le permite competir en el mercado de IA y expandirse a nuevos mercados.
- Qualcomm mantuvo su liderazgo en el mercado de procesadores para dispositivos móviles. Se enfocó en la IA y se expandió a nuevos mercados como PCs y automóviles. Es una empresa sólida que innova y se adapta al mercado.

Finalmente, 2024 fue un año devastador para Intel. Termina el año presionada por el mercado, la innovación y la Bolsa. Su CEO, Pat Gelsinger, se retira de la Presidencia.

Por otra parte, las rigurosas medidas de seguridad y restricción estadounidenses hacia China han puesto problemas de ingresos a las mismas empresas americanas de procesadores.

¿La respuesta?, una vorágine de inversiones chinas en su empresa Semiconductor Manufacturing International Corporation (SMIC). Esta empresa se convirtió por primera vez en el tercer fabricante de *chips* a nivel mundial y obtuvo 82% de sus ingresos en China, lo que documenta su dependencia local. SMIC enfrenta la posibilidad de perder acceso a la maquinaria de fabricación de *chips* desarrollada por la monopólica holandesa ASML.

Finalmente, EUA logró llevar de Asia, a su territorio, tres plantas de *chips* de Samsung, Intel y TSMC.

Grecia fue la base de Occidente con su filosofía, la construcción de sus instituciones políticas y su ficción mitológica y literaria que agregaron valores a

Occidente. Roma fue la sociedad que expandió el concepto de lo occidental, a partir de sus instituciones jurídicas y políticas. La Edad Media preservó Egipto, Roma y Grecia en sus monasterios; y el Renacimiento, con su explosión creativa, fue base de la estética occidental contemporánea. Miles de años dando tuétano al cuerpo cultural de occidente pueden palidecer con la llegada de la Inteligencia Artificial. Occidente y todo el mundo no serán los mismos. Estamos próximos al surgimiento de un renacimiento, pero digital.

III. ENTRE EL MIEDO Y LA FASCINACIÓN DIGITAL: EL SÍNDROME DE DIEGO RIVERA

La presencia de la tecnología genera temor, miedo ante su adopción. Cada generación tecnológica tiene la capacidad de generar pánico colectivo. Una mezcla de fascinación y terror perverso nos atrae tanto como nos perturba.

Entre 1930 y 1940, Diego Rivera, el enorme pintor mexicano, fue invitado a realizar distintos trabajos en Estados Unidos (EUA), pero sobresalen tres murales en San Francisco, Detroit y Nueva York.

En San Francisco, en la Escuela de Bellas Artes de California (ahora San Francisco Art Institute), Rivera pintó un mural que mostraba su fascinación e interés por la cultura industrial moderna de EUA.

En Detroit pintó los "Murales de la industria de Detroit" (1932-1933). Una hermosa serie de frescos en el Instituto de Artes de Detroit. Veintisiete paneles representan la era de la industria, específicamente de Ford Motor Company y la ciudad de Detroit. Estos murales reflejan la visión de Rivera sobre la importancia de los trabajadores en la industria automotriz, mostrando obreros de diferentes razas y su poderosa fuerza de trabajo.

Nueva York: su obra más famosa y controvertida en EUA fue "El hombre en la encrucijada", encargada para el Rockefeller Center y, posteriormente, destruida. La ciencia, la industria y el ser humano al centro de nuestra sociedad.

Cada uno de estos murales documentó la enorme admiración y los temores de Rivera a partir de la Revolución Industrial y sus efectos en la economía. Rivera mostró su fascinación, esperanza, desconfianza y aversión a la tecnología industrial. Puertos, máquinas, trenes, fábricas, aviones, autos, barcos y tecnología en todas sus dimensiones. Enormes imágenes polimórficas de la industria contrastaban con las caras de pánico, terror, sometimiento y heroísmo de ciudadanos de muchas nacionalidades. Algunos de los temores sucedieron y otros jamás se verificaron. Los sistemas de anticuerpos jurídicos actuaron y esta sociedad ha resistido.

La Inteligencia Artificial es para nuestra sociedad lo que a Rivera la Revolución Industrial, pero con una dimensión amplificada por la Internet, las redes sociales y el temor que deriva en una desconfianza recíproca en múltiples bandas: sesgos raciales, marginación, competencia económica, empleo, privacidad, ética, noticias falsas, información pública, democracia, derechos humanos, geopolítica y sistemas de defensa.

GEOPOLÍTICA

El mundo de la regulación, la política digital y las leyes en una sociedad digital son hijas directas de la geopolítica. La IA ha nacido con el nombre de la geopolítica en su frente. La geopolítica confiesa no sólo miedo sino su carencia, su ausencia profunda. La regulación digital ya venía siendo la primera frontera entre Europa, EUA y China, pero la IA despierta todos los miedos y todos los deseos al mismo tiempo. El oscuro objeto del deseo, diría Luis Buñuel.

China. En 2017, un reporte publicado en LinkedIn (Informe de talento global de Inteligencia Artificial) señalaba que, dentro de las 700 personalidades más importantes en IA, "EUA supera por más del doble a China". Según el reporte, se contaba, además, con un total de 1.9 millones de profesionales en el campo de la IA. La mayoría estaba en Estados Unidos y sólo unos 50 mil en China.

El 8 de julio de 2017 el Consejo de Estado de China lanzó el Plan de Desarrollo de la Nueva Generación de Inteligencia Artificial, donde señalaba que "el rápido desarrollo de la Inteligencia Artificial cambiará profundamente la vida social humana y cambiará el mundo. Este plan se formula… con el fin de aprovechar las principales oportunidades estratégicas para el desarrollo de la Inteligencia Artificial, ser pioneros en el desarrollo de la Inteligencia Artificial y una potencia mundial en ciencia y tecnología".

Poco después, el presidente chino Xi Jinping emite el histórico discurso donde anuncia que China será potencia en IA en 2030. Desde entonces, China ha invertido más de 150,000 millones de dólares. Hoy, China lidera la carrera de patentes de IA en términos de volumen (OMPI). Entre 2014 y 2023, más de 38,000 invenciones de IA Generativa son de China, casi seis veces más que EUA.

EUA. Domina en términos de impacto, eficiencia tecnológica e inversiones la IA. Tiene un dominio absoluto en el corazón de la IA: los procesadores. Las patentes estadounidenses se citan casi siete veces más por su calidad que las patentes chinas y concentra las mayores inversiones directas en procesadores, Data Centers y *software*.

Europa. Ha perdido competitividad respecto de EUA y China. Mario Draghi, expresidente del Banco Europeo y ex primer ministro de Italia, ha señalado que "actualmente, las empresas europeas se enfrentan a dos obstáculos cada vez más graves: una mayor competencia externa, especialmente de China", "la competitividad tiene menos que ver con los costes laborales relativos y más con el conocimiento". Europa no lidera ni en inversión, ni en patentes, ni en recursos humanos a China ni EUA, y se ha asumido en el papel de la autoridad regulatoria del mundo, un tema que el mismo Draghi señala como un problema que le quita competitividad en innovación a Europa. Ha propuesto una "pausa regulatoria".

En efecto, la guerra regulatoria que empezó todo fue la de los datos. Europa, recelosa de Estados Unidos, considera que los datos personales no están protegidos en EUA. A su vez, EUA desconfía de China por compartir sus datos e información en línea con el gobierno. Una relación digital sustentada en la desconfianza recíproca. ¿Consecuencia? Europa decidió construir una robusta legislación en diversas materias: servicios y mercados digitales, derechos humanos, protección al consumidor, propiedad intelectual, derecho de las telecomunicaciones, derecho informático, regulación *fintech*, regulación laboral, derechos digitales, protección de datos personales e Inteligencia Artificial.

Por otra parte, la regulación en materia de IA se extiende a todo el planeta y es producto de alianzas regionales, intereses específicos y de dominación económica y política. La estrategia global que busca regular la IA lo entiende y, por eso, todos quieren ser la plataforma de regulación de esta tecnología.

Los esfuerzos globales en esta materia son los que lideran la OCDE, la UNESCO, la ONUDI, la Alianza Europea de IA, la Alianza Global sobre la IA, el G7 y el grupo de Hiroshima; así como sendos documentos que ha expedido en la materia China (Ley General Reguladora de la IA y, en paralelo, la específica sobre IA Generativa) y el presidente Biden (Plan Estratégico Nacional de Investigación y Desarrollo de la IA, la Declaración de Derechos en materia de IA y la Orden Ejecutiva que instruye a las agencias federales a eliminar el sesgo en el diseño y uso de la IA y proteger al público de la discriminación algorítmica). En agosto de 2024 entró en vigor el Reglamento Europeo de Inteligencia Artificial, que tiene por objeto fomentar el desarrollo y la implantación responsable de la IA.

Europa se llena de orgullo señalando que son la primera región del mundo que cuenta con regulación para Inteligencia Artificial y que tienen el paquete más robusto de regulación digital del planeta.

La Asamblea General de la ONU adoptó una resolución histórica sobre la promoción de sistemas de Inteligencia Artificial "seguros y confiables" que beneficien el desarrollo sostenible para todos. Pidió a los Estados miembros "que se

abstengan o dejen de utilizar sistemas de IA que sean imposibles de operar con el derecho internacional de los derechos humanos".

IV. LA ERA TUMP Y EL CORPORATIVISMO DIGITAL

El presidente Trump tomó posesión de la mano de las empresas tecnológicas más poderosas del planeta, y se han reagrupado en una estructura corporativa que beneficiará a EUA en tres sectores: al interior del gobierno, frente a otros países (China, Europa y América Latina buscará invertir masivas e históricas cantidades de dinero en investigación) y desarrollo de Inteligencia Artificial. Para ese propósito, desregula el sector y presenta varias Órdenes Ejecutivas:

1. Restaura la libertad de expresión y termina con la censura federal.
2. Elimina la Orden Ejecutiva sobre Desarrollo, uso seguro, confiable y protegido de la IA.
3. Se elimina el "mandato de vehículos eléctricos".
4. Establece el Departamento de Eficiencia Gubernamental.
5. Se revoca el Consejo Presidencial de Asesores en Ciencia y Tecnología.
6. Fortalece el liderazgo estadounidense en tecnología financiera digital (criptoactivos).
7. Nombra a Brendan Carr al frente de la Comisión Federal de Comunicaciones.
8. A los miembros de la Junta de Supervisión de la Privacidad y las Libertades Civiles (PCLOB, por sus siglas en inglés) se les exige que renuncien. Es la autoridad que supervisa en Estados Unidos la privacidad.

La toma de posesión del presidente Donald Trump como presidente de Estados Unidos marca la inflexión que hacía falta en este mundo de autoritarismos. La respuesta americana está en el mundo digital.

El autoritarismo de derecha e izquierda viene con una sola capacidad: llevar cualquier conversación al extremo. No es una conversación de matices, sino de luz y sombra. Tienen urgencia por resultados y poca paciencia para el debate.

En el mundo digital, el centro del debate está en el ejercicio de los derechos humanos y políticos en este ecosistema. Pero, ¿qué pasa si el debate local se extiende al mundo? ¿Cuáles derechos humanos imperan? ¿La libertad de expresión se aplica de acuerdo a los valores europeos, estadounidenses o latinoamericanos? Europa ha defendido su derecho a defender los valores democráticos de Occidente y eso sucede en materia de derechos humanos.

Pero hay que señalar que la palabra "derechos humanos" se ha vuelto como la palabra "democracia", un concepto polisémico que se interpreta de forma distinta en todas las geografías y épocas.

Europa. Regula el mundo digital y persigue las *Big Tech* en materia fiscal, libertad de expresión, competencia económica, protección al consumidor, datos personales, derechos digitales, creó una ley de mercados y servicios digitales y ahora regula la Inteligencia Artificial. Los derechos humanos son a la europea. Les ha aplicado a las *Big Tech* multas que suman miles de millones de dólares.

Ursula Von Der Leyen, presidenta de la Comisión Europea, ha visitado EUA y a los CEOs de las *Big Tech* en varias ocasiones. Se dedicó a invitarlos a cumplir las rígidas y complejas leyes de Europa y, al mismo tiempo, a invertir allá. Las *Big Tech* sienten el asedio y persecución de Europa.

Los empresarios. La incorporación de los empresarios y poderosas empresas digitales al gobierno de Trump hablan de una visión conjunta, y ésta es la respuesta a los diversos ataques que sienten del mundo. Mark Zuckerberg anunció "la eliminación de los verificadores de hechos, porque va a regular sólo los actos ilegales"; "vamos a trabajar con el presidente Trump para hacer retroceder a los gobiernos de todo el mundo que están yendo tras de las empresas estadounidenses y presionando para censurar"; "Estados Unidos tienen las protecciones constitucionales más fuertes del mundo en materia de libertad de expresión".

Contrario sensu, "Europa tiene las leyes que institucionalizan la censura y hacen difícil construir algo innovador allá". "Vamos a trasladar nuestros equipos de confianza, seguridad y moderación de contenido fuera de California y nuestra revisión", dijo Zuckerberg.

El presidente Trump mencionó un principio de superioridad científica y tecnológica al señalar: "pusimos el conocimiento en manos de la humanidad". Si Biden creó la Ley de Ciencia, Tecnología y Chips para una superioridad en IA y cómputo, Trump anunció que "vamos a enviar a nuestros astronautas para que planten una bandera de EUA en la superficie de Marte".

La tecnología de lo diminuto y la tecnología de lo espacial. La ciencia y la tecnología como herramientas de poder económico, geopolítico y soberanía digital. Y manda un mensaje al mundo: elimina la Orden Ejecutiva sobre Desarrollo y uso seguro, confiable y protegido de la IA.

La superioridad de los valores jurídicos y derechos humanos de EUA sobre el mundo se subrayan cuando Trump señaló que, "después de años de esfuerzos ilegales e inconstitucionales para restringir la libertad de expresión, voy a firmar una Orden Ejecutiva que pone fin a toda forma de censura, para restablecer la

libertad de expresión". En el acto Zuckerberg cambia las instalaciones de Meta de California a Texas.

V. CONCLUSIONES

La IA nace como una ola tecnológica, hija de las redes de la infraestructura digital y los procesadores.

La IA es una ola tecnológica que masificará el uso de la tecnología digital y millones de nuevas aplicaciones.

La IA habilitará los derechos fundamentales de la nueva persona digital, siempre conectado en todas partes.

El riesgo inevitable es la violación o conculcación de derechos por parte de la autoridad, al sobrerregular o regular mal, y las empresas tecnológicas al facilitar la violación sistemática de derechos de los ciudadanos en múltiples materias.

El constitucionalismo digital es una aproximación epistémica fundamental para promover la adopción, ejercicio de la tecnología digital y protección de los derechos fundamentales en un entorno de protección y respeto a la ley, al ciudadano y su derecho al uso y disfrute de estas tecnologías.

El análisis del fenómeno digital y la IA requiere una aproximación multidisciplinaria, de materias jurídicas, económicas y tecnológicas. Estamos frente a un fenómeno complejo que debe verse así, para aproximarse a él y comprenderlo mejor.

La gobernanza de lo digital y la IA comienza por comprender que son fenómenos que nacen en un entorno geopolítico. Europa se siente amenazada por China y EUA. La Unión Americana siente desconfianza en China y se siente agredido por Europa. China siente desconfianza de Occidente. Estamos frente a una cultura de la desconfianza recíproca.

El fenómeno de la IA se sostiene con inversiones económicas, nunca antes vistas en nuestra civilización, y se presentan en un entorno de libre mercado extremo (EUA) y un capitalismo de Estado (China).

La sujeción de la tecnología digital en todas sus expresiones, incluida la IA, a la ley, será el mayor reto de nuestra sociedad, particularmente por la toma de decisiones aceleradas en todas las áreas del conocimiento humano y su impacto en los derechos humanos y la economía.

La IA avanza mucho más rápido de lo que se esperaba. La Ley de Moore (Gordon Moore), que se sostiene bajo el principio de que cada dos años se duplica la capacidad de procesamiento de datos, está siendo literalmente

destrozada y se comienza a superar casi cada año.

La regulación de la IA y la tecnología digital nos pone frente a un choque de principios jurídicos y valores dentro de Occidente y en el mundo. Europa regula la tecnología digital y la IA bajo principios jurídicos y valores que considera universales, bajo el concepto de lo occidental. Eso está cuestionado hoy día por EUA, que considera que los principios básicos de la democracia y los derechos humanos se encuentran mejor interpretados en EUA, e incluso, dentro de EUA, dentro del Estado de Texas.

En consecuencia, el derecho internacional se enfrentará a:

La necesidad de observar y generar prospectiva regulatoria, antes de generar marcos legales que inhiban la innovación o conculquen derechos fundamentales.

La necesidad de observar la rápida evolución de la IA, para comprender mejor su impacto multidimensional en la sociedad.

Al reto de un marco jurídico internacional fragmentado y con distintas interpretaciones sobre lo que son los derechos fundamentales, la regulación, la política pública y dominados por poderosos intereses económicos y geopolíticos. En suma, estamos frente al reto de la mayor conversación global de nuestra civilización.

CAPÍTULO DOS

INTELIGENCIA ARTIFICIAL. LÍNEAS GENERALES DE UN NUEVO CONOCIMIENTO

PASADO, PRESENTE Y FUTURO DE LA INTELIGENCIA ARTIFICIAL

SUMARIO

I. Introducción **II.** Precursores de la inteligencia artificial **III.** El nacimiento de una disciplina **IV.** Los primeros éxitos de la inteligencia artificial **V.** El primer invierno de la IA (1974-1980) **VI.** El resurgimiento de la IA (1980-1987) **VII.** El segundo invierno de la IA (1987-1993) **VIII.** Los resultados prometidos (1993-2011) **IX.** El uso de grandes cantidades de datos (2011-fecha) **X.** El Futuro de la inteligencia artificial **XI.** El miedo a la inteligencia artificial **XII.** Conclusiones **XIII.** Bibliografía

> CARLOS A. COELLO COELLO
Miembro del Colegio Nacional. Investigador en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN) en Ciudad de México y Profesor Visitante en el Basque Center for Applied Mathematics en España

I. INTRODUCCIÓN

Actualmente, la Inteligencia Artificial (IA) goza de una enorme popularidad y resulta claro que ya comienza a tener un enorme impacto en nuestra vida diaria y se espera que dicho impacto sea todavía mayor en los próximos años.

Al igual que como ha ocurrido con otras tecnologías, la IA ha despertado un asombro que después se ha convertido en miedo, principalmente por el desconocimiento que existe en torno a sus alcances y limitantes. Adicionalmente, la ciencia ficción ha alimentado, durante muchos años, el temor a la IA a través de un gran número de relatos, películas y series televisivas.

En este capítulo abordaremos la evolución de la IA, desde sus remotos orígenes (en la década de 1950) hasta nuestros días. A lo largo de este recorrido, veremos cómo han ido evolucionando los algoritmos de IA, que pasaron de realizar tareas en modelos muy simples y reducidos, a ganarle al campeón del mundo en ajedrez o a sostener conversaciones realistas con un humano. En este recorrido, transitaremos de la denominada IA simbólica a la IA subsimbólica que utiliza grandes cantidades de datos, que es la predominante hoy en día.

Es claro que la IA, entre otras cosas, ha revolucionado a la ciencia misma y la forma de hacer investigación. Esto fue reconocido en 2024 por el comité del Premio Nobel. Pero además de los avances científicos que la IA está impulsando, también está cambiando nuestra vida diaria con aplicaciones cada vez más poderosas que nos permiten potenciar nuestra creatividad.

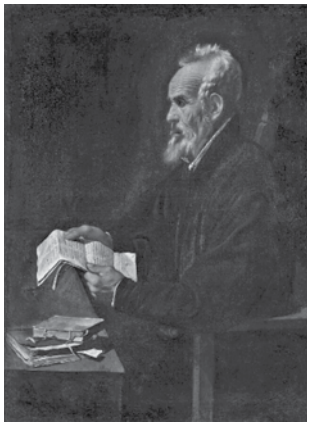
Los bebés que nacen hoy en día, crecerán con algoritmos de IA cada vez más poderosos y sofisticados, y seguramente serán capaces de asimilar esta nueva tecnología con mayor facilidad que los adultos, al igual que ha sucedido con los teléfonos y los televisores inteligentes.

Pero quizás la lección más importante por aprender es que, nos guste o no, la IA llegó para quedarse y por ello, será importante que intentemos adaptarnos a ella, aprendiendo al menos lo más básico sobre su funcionamiento. Eso evitará atemorizarnos con cualquier nota alarmista que leamos en las redes sociales y nos permitirá hacer un uso más racional y adecuado de esta tecnología. Este capítulo busca contribuir un poco a ese conocimiento básico tan necesario que debemos tener de la IA y de sus alcances.

II. PRECURSORES DE LA INTELIGENCIA ARTIFICIAL

La inteligencia es algo que ha fascinado a la humanidad durante un largo tiempo. La premisa fundamental en la que se basa la inteligencia artificial es que el proceso de generar pensamientos se puede mecanizar [Russell y Norvig, 2021]. El estudio del razonamiento mecánico (o formal) es muy antiguo, y está documentado al menos desde unos 1000 años antes de Cristo en China, India y Grecia. El filósofo español Ramón Llull (ver figura 1) diseñó y construyó (en el siglo XIII) una máquina lógica en la que las teorías, los sujetos y los predicados teológicos se representaban mediante figuras geométricas que él consideraba "perfectas" (p.ej., círculos, cuadrados y triángulos). Usando palancas, manivelas y un volante, era posible demostrar con esta máquina la falsedad o certeza de un postulado [Mata, 2006]. El trabajo de Llull tuvo una enorme influencia en el de Gottfried Leibniz, que llegó a especular (en el siglo XVII) sobre la existencia futura de un lenguaje universal para expresar los razonamientos. Leibniz creía que en el futuro, los filósofos ya no tendrían que debatir, sino que sería posible demostrar matemáticamente si sus propuestas eran o no correctas.

FIGURA 1: Ramón Llull.

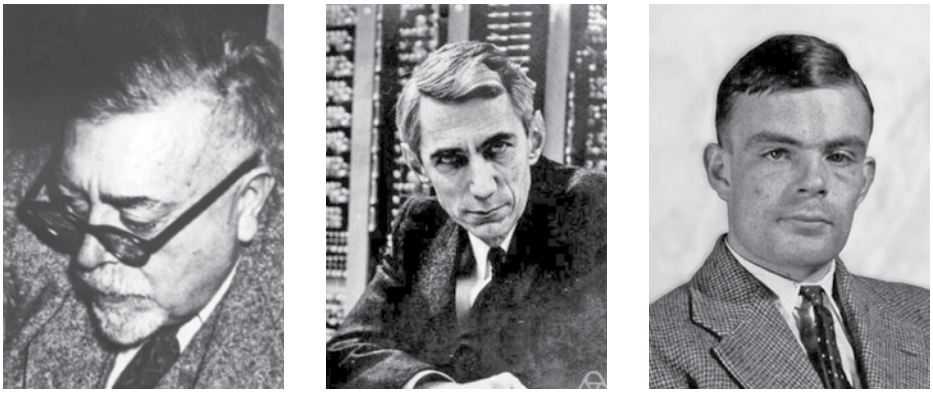


FUENTE: Wikipedia (imagen en el dominio público)

Todas estas ideas tuvieron una enorme influencia en el desarrollo de la "hipótesis de los símbolos físicos" de Allen Newell y Herbert Simon, que fue la base de la denominada inteligencia artificial simbólica que fue la corriente que predominó en los orígenes de esta disciplina.

Sin embargo, históricamente, suelen considerarse a tres personajes como los precursores principales de la IA: (1) Norbert Wiener, (2) Claude

FIGURA 2: De izquierda a derecha: Norbert Wiener, Claude Shannon y Alan Turing.



FUENTE: Wikipedia (imágenes en el dominio público).

Shannon y (3) Alan Turing. A continuación, se describirán brevemente las contribuciones principales (relacionadas con la Inteligencia Artificial) realizadas por cada uno de ellos.

Norbert Wiener fue un matemático norteamericano que propuso en la década de 1930 (junto con el mexicano Arturo Rosenblueth) una disciplina a la que denominó cibernética [Wiener, 1948], la cual estudia (de forma transdisciplinaria) los procesos circulares tales como los sistemas de retroalimentación en los que las salidas actúan también como entradas. En sus orígenes, esta disciplina se aplicó en diversas áreas, incluyendo procesos biológicos, cognitivos, mecánicos y hasta sociales.

Claude Shannon fue un ingeniero norteamericano que publicó un artículo [Shannon, 1948] en el que originó una disciplina que después sería denominada "teoría de la información", la cual consiste en una teoría matemática de la comunicación. En este artículo, Shannon presenta los conceptos básicos de las comunicaciones digitales y desarrolla además conceptos como la entropía de la información y la redundancia. Adicionalmente, Shannon introdujo el término *bit* (que después se atribuiría a John Tukey) que usó para denominar a una unidad de información.

Alan Turing fue un matemático inglés que publicó un artículo [Turing, 1950] que después se volvería muy famoso, en el cual propone una definición operativa de inteligencia. La propuesta de Turing es muy simple. Plantea una prueba (que hoy se denomina "prueba de Turing") que consiste en tener a una persona frente a la terminal de una computadora y poner en cuartos separados a otro humano y a una computadora. Esta persona estará interactuando a través de la

computadora sin saber quién le respondió (si el otro humano o la computadora). En el momento en que el humano no pueda distinguir quién le respondió puesto que las respuestas de la computadora son indistinguibles de las de un humano, se podrá decir (según Turing) que esa computadora es inteligente.

Sin embargo, Turing es más recordado por un trabajo en el que definió un modelo matemático de una computadora universal (hoy en día, ese modelo se conoce como "máquina universal de Turing"). Turing definió lo que es computable (es decir, los problemas que pueden resolverse en una computadora) y lo que no, originado lo que hoy se conoce como "teoría de la computación".

En la década de 1940 se llevaron a cabo avances importantes en torno a esta disciplina que todavía no tenía nombre. De entre ellos, destaca el artículo titulado "*A logical calculus of the ideas immanent in nervous activity*" [McCulloch y Pitts, 1943]. En este artículo, Warren McCulloch y Walter Pitts proponen un modelo de neuronas artificiales en el cual cada neurona está caracterizada por estar, ya sea "encendida" o "apagada". En este modelo, una neurona se "enciende" en respuesta a la estimulación de un número suficiente de neuronas vecinas. En este artículo, McCulloch y Pitts logran demostrar que cualquier función computable puede ser modelada por alguna red de neuronas conectadas y que todos los operadores lógicos (p.ej., AND, OR, NOT) pueden ser implementados mediante estructuras simples en la red. También sugirieron que redes de neuronas que estuvieren configuradas adecuadamente serían capaces de aprender. Para desarrollar este trabajo, McCulloch y Pitts se basaron en tres fuentes principales: (1) su conocimiento sobre la fisiología básica y el funcionamiento de las neuronas en el cerebro, (2) un análisis formal de lógica proposicional [Whitehead y Russell, 1910] y (3) la nueva teoría de la computación propuesta por Alan Turing. Este artículo que suele considerarse como el primero sobre IA en la historia [Rusell y Norvig, 2021], constituye una contribución seminal para un área de investigación que se volvería muy popular muchos años después: las redes neuronales artificiales.

FIGURA 3: Marvin Minsky.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

En la misma década de 1940, Donald Hebb [1949] creó una hipótesis sobre el aprendizaje basada en el mecanismo de la plasticidad neuronal. En este trabajo, propone una regla

muy simple de actualización para modificar la fortaleza de la conexión entre neuronas que hoy se conoce como aprendizaje de Hebb y que se usa todavía en la actualidad para el denominado denominado aprendizaje no supervisado.

En 1951, influenciado por el artículo de McCulloch y Pitts antes mencionado, Marvin Minsky, que en ese entonces cursaba un doctorado en matemáticas en la Universidad de Princeton), construyó, con la ayuda de Dean S. Edmonds (quien cursaba un doctorado en física en la misma universidad) el primer modelo físico de una red neuronal. Este dispositivo, que fue denominado *Stochastic Neural Analog Reinforcement Computer* (SNARC) usaba 300 bulbos para simular una red de 40 neuronas que usaban aprendizaje de Hebb. SNARC fue capaz de hacer que unos ratones encontraran la salida de un laberinto mediante una simulación que se visualizaba como un "arreglo de luces" (este es un problema que planteó originalmente Claude Shannon). Un circuito se reforzaba cada vez que uno de los ratones simulados llegaba a la meta. SNARC sorprendió a sus creadores, quienes luego dirían que los ratones se ayudaban entre sí en la simulación, de tal forma que si uno encontraba una buena ruta, los demás tendían a seguirlo. En su tesis doctoral, Minsky estudió el cómputo universal en las redes neuronales, lo cual le acarreó algunos problemas. Su comité doctoral veía con escepticismo su trabajo de investigación, pues no estaban seguros de que pudiese considerarse como parte de las matemáticas. Pero se dice que John von Neumann fue quien resolvió la situación, pues al preguntársele si creía que este trabajo era parte de las matemáticas, respondió "si no lo es ahora, lo será algún día". Con los años, Marvin Minsky se convirtió en una de las figuras más influyentes de la IA en el mundo.

III. EL NACIMIENTO DE UNA DISCIPLINA

En la década de 1950, la Universidad de Princeton contaba con otra persona que se volvería una figura muy prominente dentro de la IA: John McCarthy. Después de doctorarse en matemáticas en la Universidad de Princeton en 1951, McCarthy trabajó ahí como instructor durante dos años. Posteriormente, y tras un breve período en la Universidad Stanford, fue contratado como Profesor Asistente en el Colegio Dartmouth.

FIGURA 4: John McCarthy.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

McCarthy convenció a Marvin Minsky, Claude Shannon y Nathaniel Rochester para que lo ayudaran a reunir a un grupo de investigadores norteamericanos interesados en teoría de autómatas, redes neuronales y el estudio de la inteligencia. Estos esfuerzos condujeron a la realización de un taller de dos meses de duración, el cual se llevó a cabo en el Colegio Dartmouth, en el verano de 1956. Este evento marca oficialmente el nacimiento de la Inteligencia Artificial, que es un término que acuñó John McCarthy en la propuesta que redactó para conseguir fondos para el taller, en la cual decía lo siguiente [Russell y Norvig, 2021]:

> Proponemos la realización de un taller de 2 meses de duración en el Colegio Dartmouth, en Hanover, New Hampshire, en el que 10 personas estudiarán la inteligencia artificial. La idea es proceder a partir de la conjetura de que todo aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, ser descrita de forma tan precisa que es posible hacer que una máquina la simule.

Se dice que McCarthy evitó usar el término "computadora" al nombrar la nueva disciplina, pues buscaba distanciarse del trabajo de Norbert Wiener en torno a la cibernética, en el cual se usaban computadoras analógicas y no digitales como las que McCarthy tenía en mente utilizar [Russell y Norvig, 2021]. La conjetura indicada en la propuesta, sería llamada después "la hipótesis de los sistemas de símbolos físicos" y constituiría la base de la denominada "inteligencia artificial simbólica". Esta hipótesis sería extendida, reforzada y bautizada por Allen Newell y Herbert Simon en su discurso de aceptación del Premio Turing, en 1976 [Newell y Simon, 1976].

Los invitados a este taller crearían programas muy importantes para el desarrollo inicial de la inteligencia artificial: Arthur Samuel, Nathaniel Rochester, Ray Solomonoff, Oliver Selfridge, Trenchard More, Allen Newell y Herbert A. Simon. Pero los dos últimos fueron las estrellas del taller, pues habían desarrollado un programa llamado *Logic Theorist* [McCorduck, 2004], el cual fue usado para demostrar la mayor parte de los teoremas del capítulo 2 del *Principia Mathematica* [Whitehead y Russell, 1910]. El *Logic Theorist* fue capaz de encontrar demostraciones más simples que las del libro.

IV. LOS PRIMEROS ÉXITOS DE LA INTELIGENCIA ARTIFICIAL

El inicio de esta nueva disciplina estuvo marcado por diversos éxitos que hicieron que los investigadores asociados a ellos desbordaran entusiasmo. Muchos de los proyectos iniciales de esta disciplina fueron financiados en Estados Unidos por la Agencia de Proyectos de Investigación Avanzados (*Defense Advanced Research Projects Agency*, o DARPA).

Después del rotundo éxito del *Logic Theorist*, Allen Newell y Herbert Simon propusieron el *General Problem Solver* (GPS), que estaba diseñado para imitar los protocolos usados por los humanos para resolver problemas [McCorduck, 2004]. El GPS podía resolver un conjunto muy limitado de problemas, pero en ellos, este programa consideraba submetas y acciones posibles, de forma similar a como lo hacemos los humanos. El éxito del GPS y de programas posteriores que operaban como modelos cognitivos, llevaron a Newell y Simon a formular la famosa hipótesis de los símbolos físicos que dice [Newell y Simon, 1976]: "un sistema de símbolos físicos tiene los medios necesarios y suficientes para producir acciones inteligentes generales".

Esta hipótesis fue la base de la denominada Inteligencia Artificial simbólica y sería desafiada a lo largo de los años. Con ella, la idea fundamental detrás de la IA era que cualquier sistema que exhibiera inteligencia, debía operar mediante la manipulación de estructuras de datos compuestas de símbolos [Russell y Norvig, 2021].

En IBM, Nathaniel Rochester y sus colegas, produjeron avances importantes para la Inteligencia Artificial. Por ejemplo, Herbert Gelernter [1959] construyó el *Geometry Theorem Prover*, que era capaz de resolver teoremas de geometría Euclideana que muchos estudiantes de matemáticas consideraban complicados.

En 1952, Arthur Samuel (también de IBM) escribió una serie de programas para jugar damas. Samuel estaba convencido de que enseñar a una computadora a jugar era una tarea muy fructífera que conduciría al desarrollo de tácticas para resolver problemas más generales. Para jugar damas, Samuel usó árboles de búsqueda de las posiciones del tablero que eran alcanzables desde el estado actual del juego. Como las computadoras que usó tenían muy poca memoria, implementó una técnica que después se conocería como "poda alfa-beta" [Sutton y Barto, 2018]. Así mismo, en vez de buscar cada ruta hasta llegar al final de una partida, Samuel desarrolló una función que daba un puntaje a la posición actual del tablero en cualquier momento dado, y se basaba en ella para realizar movimientos usando una estrategia minimax, que buscaba maximizar la posición

del jugador que movía, mientras minimizaba el puntaje del contrincante. Samuel también propuso en 1955 diversos mecanismos que hacían que su programa mejorara su nivel de juego, jugando miles de partidas contra sí mismo. Y, de hecho, se atribuye a Samuel el haber acuñado el término "aprendizaje máquina" (*machine learning*) [Samuel, 1959].

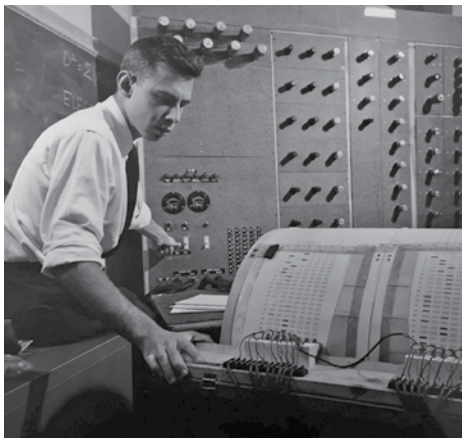
John McCarthy dejó su plaza en el Colegio Dartmouth para irse al Instituto Tecnológico de Massachusetts (MIT por sus siglas en inglés), en donde realizó tres contribuciones seminales a la IA en un solo año (1958). La primera fue el desarrollo de un lenguaje de programación de alto nivel, llamado LISP, que sería utilizado durante muchos años para desarrollar aplicaciones de IA en todo el mundo. La segunda contribución fue su invención (junto con otros investigadores del MIT) del tiempo compartido, que fue un mecanismo que permitió mejorar el acceso a los escasos y costosos recursos de cómputo disponibles en esa época. Y la tercera y última de estas contribuciones fue un artículo titulado *Programs with Common Sense* [McCarthy, 1958], en el cual describió el *Advice Taker*, un programa hipotético que podía verse como el primer sistema completo de IA. Al igual que el *Logic Theorist* y el *Geometry Theorem Prover*, el programa de McCarthy fue diseñado para usar conocimiento para buscar soluciones a los problemas. Sin embargo, a diferencia de los otros sistemas de la época, el *Advice Taker* buscaba conjuntar conocimiento general del mundo. McCarthy mostró, por ejemplo, cómo, a través de axiomas simples, su sistema podía generar un plan para conducir un auto al aeropuerto. Adicionalmente, McCarthy mostró que su programa estaba diseñado para aceptar nuevos axiomas durante su operación, lo cual le permitía realizar acciones en áreas nuevas sin tener que ser reprogramado. De tal forma, el *Advice Taker* ilustraba de manera fiel la hipótesis de los símbolos físicos, pues evidenciaba la importancia de contar con una representación formal del conocimiento y del uso de procesos deductivos para manipular dicho conocimiento.

1958 fue también el año en que Marvin Minsky ingresó a trabajar en el MIT. Aunque Minsky colaboró inicialmente con McCarthy (ambos fundaron el Laboratorio de Inteligencia Artificial del MIT en 1959), acabaron por separarse, pues McCarthy tenía interés en desarrollar modelos de representación y razonamiento basados en lógica formal, mientras que Minsky estaba más interesado en hacer que los programas funcionaran en aplicaciones prácticas. En 1961, James Slagle escribió (en LISP) el primer programa de integración simbólica, llamado SAINT (*Symbolic Automatic INTegrator*), como parte de su tesis doctoral [Slagle, 1961], supervisada por Marvin Minsky en el MIT. En 1964, Daniel G. Bobrow presentó en su tesis doctoral del MIT un programa, escrito en LISP,

llamado STUDENT, que era capaz de resolver problemas de álgebra a nivel preparatoria a partir de su descripción en lenguaje natural [Russell y Norvig, 2021].

En 1957, Frank Rosenblatt (ver figura 5), en la Universidad Cornell, produjo la primera implementación de un perceptrón multi-capa, el cual se demostró públicamente el 23 de junio de 1960. Rosenblatt describió los detalles de su perceptrón en un artículo de 1958 [Rosenblatt, 1958]. Su modelo consistía de una capa de entrada, una capa oculta (que no podía aprender) con pesos aleatorios y una capa de salida con conexiones que permitían aprender. Aunque el objetivo de Rosenblatt era producir una implementación en hardware de su perceptrón, inicialmente lo simuló en software en una computadora IBM 704, del Laboratorio Aeronáutico de la Universidad Cornell. Rosenblatt extendió su perceptrón en diversos artículos y en un libro [Rosenblatt, 1962], llegando a demostrar cuatro teoremas que sentaron las bases teóricas de su propuesta. En su libro, Rosenblatt incluyó diversos experimentos con perceptrones de hasta dos capas, los cuales entrenó usando "errores propagados hacia atrás". Sin embargo, no utilizó el algoritmo que hoy se conoce como de "propagación hacia atrás" (*backpropagation*) y, de hecho, Rosenblatt no contaba con un método general para entrenar perceptrones de capas múltiples. Aunque el trabajo de Rosenblatt fue ampliamente reconocido y atrajo fuertemente el interés de la prensa, posteriormente, este interés decreció debido a un artículo publicado por Marvin Minsky y Seymour Papert en 1969, en el cual criticaban las limitantes del perceptrón, sin aclarar que sus resultados se obtuvieron con una versión del perceptrón con limitantes en sus entradas [Kirdin *et al.*, 2022].

FIGURA 5: Frank Rosenblatt.



FUENTE: Cornell University division of rare and manuscript collections.

En la misma dirección del trabajo de Rosenblatt, se produjeron otros avances importantes para las redes neuronales en la década de 1960. Por ejemplo, Winograd y Cowan (1963) mostraron cómo un gran número de elementos podían representar colectivamente un concepto individual, acompañados de un incremento en la robustez y el paralelismo. Bernard Widrow propuso,

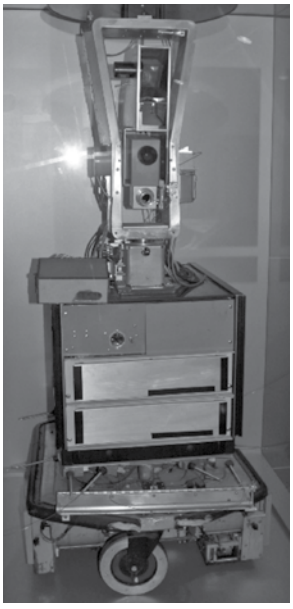
junto con su estudiante doctoral Marcian Edward (Ted) Hoff, un algoritmo adaptativo basado en un filtro de mínimos cuadrados (Widrow y Hoff, 1960). Este algoritmo permitió el desarrollo de las redes neuronales ADALINE (*ADAptive Linear NEuron*) y MADALINE (*Many ADALINE*), y posteriormente condujo al desarrollo del algoritmo de propagación hacia atrás (*backpropagation*).

FIGURA 6: John Alan Robinson.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

En 1963, McCarthy fundó el laboratorio de Inteligencia Artificial de la Universidad de Stanford. Su plan de usar lógica formal para construir una versión avanzada del *Advice Taker* fue catapultado por el método de resolución que propuso John Alan Robinson (ver figura 6) en 1965 (Robinson, 1965), el cual consistía en un algoritmo completo para demostrar teoremas basado en lógica de primer orden. La investigación realizada en la Universidad de Stanford de esa época enfatizaba los métodos de propósito general para el razonamiento lógico. Uno de los proyectos más emblemáticos de este grupo de investigación fue el robot *Shakey* (ver figura 7). Este robot fue el primero en ser controlado usando técnicas de IA y fue el primero en demostrar la integración completa del razonamiento lógico y la acción física.

FIGURA 7: El robot Shakey.



FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons)*.

El resto de la década de 1960 fue muy fructífera en descubrimientos que tendrían un enorme impacto en la IA varios años después. Por ejemplo, Alexsey Ivakhnenko y Valentin Lapa desarrollaron el primer algoritmo de aprendizaje profundo para perceptrones multi-capa en la Unión Soviética [Ivakhnenko y Lapa, 1967]. Así mismo, Lofti Zadeh publicó su artículo seminal sobre la lógica difusa [Zadeh, 1965] y se propusieron las versiones originales de los tres tipos principales de algoritmos

evolutivos: los algoritmos genéticos [Holland, 1962], la programación evolutiva [Fogel *et al.*, 1966] y las estrategias evolutivas [Rechenberg, 1965; Schwefel, 1965]. Todas estas contribuciones originarían una disciplina que se denominaría años después "inteligencia computacional" [Kacprzyk y Pedrycz, 2015].

Entre 1964 y 1966, Joseph Weizenbaum (ver figura 8), que era profesor en el Instituto Tecnológico de Massachusetts, desarrolló un programa (escrito en un lenguaje de programación llamado SLIP, que fue creado por el mismo Weizenbaum) al que nombró ELIZA [Weizenbaum, 1966]. Este programa simulaba una conversación usando un método muy simple de reconocimiento de palabras clave en una oración. Aunque daba a sus usuarios la sensación de entender lo que le decían, el programa no contaba siquiera con una representación del texto que recibía como entrada. ELIZA usaba diferentes scripts. El más famoso de ellos se llamaba DOCTOR y simulaba un sicólogo de la escuela Rogeriana (llamada así en honor a su creador, Carl Rogers), pues en este caso, el terapeuta repite las palabras de sus pacientes.

FIGURA 8: Joseph Weizenbaum.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

FIGURA 9: Shunichi Amari.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

ELIZA fue el primer chatbot de la historia y causó una gran preocupación en Weizenbaum por la forma en la que muchas personas solían contarle cosas muy íntimas. Esta preocupación acabó por convertir a Weizenbaum en uno de los mayores detractores de la Inteligencia Artificial. En 1976 publicó un libro titulado "*Computer Power and Human Reason: From Judgement to Calculation*" [Weizenbaum, 1976], en el cual indica que si bien la IA podría volverse una realidad, nunca debiéramos permitirle a las computadoras tomar decisiones importantes porque carecen de cualidades humanas importantes tales como la compasión y la sabiduría. La idea central de su libro es que existe una diferencia fundamental entre decidir y elegir. Para él, decidir es algo que se puede

programar, pero elegir es algo que debe hacerse mediante juicios y no solo mediante cálculos.

En 1967, Shunichi Amari (ver figura 9) reportó la primera red neuronal multi-capa en ser entrenada con un método de gradiente estocástico [Amari, 1967]. Esta red neuronal fue capaz de clasificar patrones separables no linealmente (algo que el perceptrón original de Rosenblatt no podía hacer).

En 1968, Richard Greenblatt desarrolló, en el Instituto Tecnológico de Massachusetts, el primer programa de computadora para jugar ajedrez a nivel de torneo: *Mac Hack*. Este programa fue también el primero en derrotar a un contrincante humano y en competir en un torneo de ajedrez de humanos.

El algoritmo de propagación hacia atrás (*backpropagation*) se desarrolló de forma independiente a principios de la década de 1970. El primero lo propuso el finlandés Seppo Linnainmaa en su tesis de maestría [Linnainmaa, 1970; Linnainmaa, 1976]. Posteriormente, Paul Werbos lo desarrolló de forma independiente (es decir, sin saber del trabajo de Linnainmaa) en 1974 [Anderson y Rosenfeld, 2000] pero debido a diversas dificultades, lo logró publicar hasta ocho años después [Werbos, 1982]. Sin embargo, fue un artículo posterior [Rumelhart *et al.*, 1986] el que popularizó el uso del algoritmo de propagación hacia atrás.

V. EL PRIMER INVIERNO DE LA IA (1974-1980)

Los investigadores más destacados en los orígenes de la IA desbordaron optimismo y realizaron predicciones muy ambiciosas. Por ejemplo, Herbert Simon dijo en 1957 que el campeón mundial de ajedrez sería derrotado por una computadora en 10 años. Su predicción eventualmente se cumplió, pero tuvieron que transcurrir 40 años para ello.

Este optimismo produjo el "primer invierno de la Inteligencia Artificial" de 1974 a 1980. Se denomina así a este período, porque durante él, se redujo considerablemente el apoyo económico de las agencias principales de financiamiento de la IA debido, sobre todo, a que los resultados prometidos en la década de 1950 no se volvieron una realidad [Russell y Norvig, 2021].

Las predicciones tan optimistas que se realizaron en la década de 1950 no se cumplieron debido a varias razones. La primera de ellas fue que estas predicciones se basaron en resultados obtenidos a través de ejemplos muy sencillos y en una escala muy pequeña. Cuando estos primeros sistemas basados en Inteligencia Artificial simbólica intentaron aplicarse a problemas

más complejos, fracasaron estruendosamente. Esto se debió principalmente a que la mayoría de los primeros programas de IA que se desarrollaron no sabían nada sobre el dominio en que se aplicaban. Estos programas realizaban simples manipulaciones sintácticas.

Un ejemplo de estas limitaciones ocurrió en el área de traducción automática (llamada *machine translation* en inglés). En 1957, tras el lanzamiento exitoso del satélite ruso Sputnik, el *National Research Council* de Estados Unidos intentó acelerar la traducción de artículos científicos escritos en ruso. Para ello, recurrieron a expertos en IA que usaron técnicas basadas en simples transformaciones sintácticas realizadas a partir de las gramáticas del ruso y el inglés y usando un simple reemplazo de palabras contenidas en un diccionario electrónico. Evidentemente, este enfoque era incorrecto, porque la traducción de textos requiere conocer el contexto de las oraciones, pues una misma palabra puede tener varios significados distintos.

Un reporte elaborado en 1966 por un comité de asesores del *National Research Council* indicó que no era posible utilizar las técnicas existentes de IA para automatizar la traducción de textos científicos del ruso al inglés. Este reporte provocó la cancelación de todos los fondos del gobierno de Estados Unidos para esa área [Russell y Norvig, 2021].

Antes de que se desarrollara la teoría de la complejidad dentro de las ciencias de la computación, se creía que "escalar" un programa a problemas más grandes era cuestión simplemente de tener hardware más rápido y memorias más grandes. Los primeros programas de IA utilizaban técnicas semi-exhaustivas, pues solían probar muchas combinaciones para elegir la correcta. Este tipo de técnicas solo funcionan adecuadamente en problemas pequeños, pero no pueden usarse en problemas grandes, porque en ellos, la cantidad de opciones disponibles crece exponencialmente, haciendo imposible la aplicación de búsquedas semi-exhaustivas. A esto se le conoce como la "maldición de la dimensionalidad".

El tercer problema fue el tipo de estructuras que se utilizaron para las primeras técnicas usadas para generar comportamientos inteligentes. Muchas de ellas tenían limitaciones inherentes a la estructura en sí, las cuales no dependían de la escala del problema. Ese era el caso, por ejemplo, del perceptrón de una capa de Rosenblatt que no podía clasificar correctamente clases que no pudieran dividirse usando una línea recta. Aunque este problema no lo tienen los perceptrones multi-capa, éstos fueron poco populares en los orígenes de las redes neuronales, debido a que su entrenamiento era mucho más complejo.

VI. EL RESURGIMIENTO DE LA IA (1980-1987)

En el período de 1980 a 1987, ocurrieron tres acontecimientos que permitieron un resurgimiento de la investigación en IA:

1. El éxito de los sistemas expertos desarrollados por Edward Feigenbaum.
2. El proyecto de la "quinta generación" desarrollado por el sector gubernamental y el sector industrial de Japón.
3. John Hopfield y David E. Rumerhart logran revivir el interés en el conexionismo (o sea, las redes neuronales), el cual regresaría para quedarse.

A continuación, discutiremos brevemente cada uno de estos eventos.

FIGURA 10: Edward Feigenbaum.

FUENTE: Wikipedia (disponible en el dominio público).

LOS SISTEMAS EXPERTOS

En 1965, Edward Feigenbaum (ver figura 10), un investigador de la Universidad Stanford, se interesó en un problema que le propuso el genetista Joshua Lederberg (ganador del Premio Nobel): inducir la estructura tridimensional de compuestos químicos a partir de su espectrograma de masas. Feigenbaum y sus colaboradores trabajaron durante 10 años (hasta 1975) en un sistema llamado *Heuristic* DENDRAL (después el nombre se recortó a DENDRAL) el cual se utilizó para inferir estructuras geométricas de compuestos químicos complejos, dada su fórmula química y sus datos del espectrograma de masas. DENDRAL descubrió algunas estructuras previamente desconocidas, y estos descubrimientos se publicaron en una serie de artículos que aparecieron en el *Journal of the American Chemical Society*.

DENDRAL usaba reglas, obtenidas de los químicos, relacionadas con la forma en la que un espectrómetro de masas fragmentaba compuestos y los volvía sub-estructuras. A partir del conocimiento de estas sub-estructuras, DENDRAL era capaz de deducir la estructura más factible del compuesto. Posteriormente, una nueva versión del programa, llamada Meta-DENDRAL, fue capaz incluso de deducir automáticamente nuevas reglas a partir de datos químicos. Estas

nuevas reglas permitían mejorar el desempeño de DENDRAL [Buchanan y Feigenbaum, 1978]. Este trabajo convenció a Feigenbaum de la importancia de dotar a los programas de computadora de conocimiento en forma de reglas y procedimientos, a fin de guiarlos para la solución de problemas. Posteriormente, Feigenbaum llamaría "sistema experto" al sistema de reglas utilizado por DENDRAL.

Los sistemas expertos fueron la primera herramienta de IA que se aplicó de manera exitosa en problemas del mundo real. Feigenbaum desarrolló posteriormente sistemas expertos para otras áreas, incluyendo medicina (MYCIN) y genética molecular (MOLGEN).

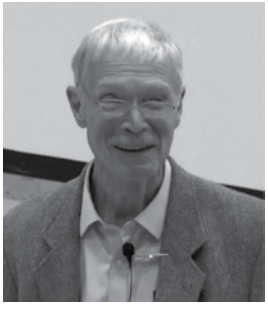
EL PROYECTO DE LA QUINTA GENERACIÓN

En 1981, el gobierno de Japón en colaboración con su sector industrial, planteó el proyecto denominado Sistemas de Cómputo de Quinta Generación. El plan original establecía una duración de 10 años y era muy ambicioso, pero poco realista. La propuesta original fue modificada en 1982 por otra con objetivos más realistas.

Entre otras cosas, este proyecto planteaba que los sistemas de cómputo de quinta generación contarían con: (1) mecanismos para resolución de problemas y para realizar inferencia, (2) manejo de enormes bases de conocimiento e (3) interfaces inteligentes que usaran lenguaje natural tanto de forma impresa como hablada. Evidentemente, esto requeriría un enorme poder de cómputo y se planteó el uso de hardware para el manejo de bases de datos paralelas y software para el manejo de bases de conocimiento. También se usaría hardware para inferencia en paralelo, manejo de tipos de datos abstractos y máquinas de flujo de datos. Adicionalmente, habría hardware especializado para reconocimiento de voz e interfaces que funcionarían usando lenguaje natural.

En vez de elegir LISP para este ambicioso proyecto, los japoneses decidieron adoptar PROLOG, que es un lenguaje de programación que fue desarrollado en Francia por Alain Colmerauer y Philippe Roussel a inicios de la década de 1970.

FIGURA 11: John J. Hopfield.



FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

Aunque en sus primeros 10 años, el proyecto de la quinta generación estaba lejos del cumplimiento de sus metas originales [Feigenbaum y Shrobe, 1993], su planteamiento atrajo fuertemente la atención de Estados Unidos y despertó su temor de que Japón se adueñara de la industria de cómputo y de los sistemas inteligentes. Para evitarlo, Estados Unidos decidió invertir enormes cantidades de dinero para financiar investigación en torno al paralelismo a principios de la década de 1980. Este financiamiento contribuyó de forma muy importante al resurgimiento de la IA.

EL REGRESO DE LAS REDES NEURONALES

En 1982, el físico John Hopfield (ver figura 11) publicó un artículo [Hopfield, 1982] en el que propone un tipo de red neuronal recurrente que ahora lleva su nombre (también se le llama memoria asociativa) en la cual se modela a las neuronas para que interactúen de acuerdo a la dinámica de giro de los sistemas físicos tales como los materiales magnéticos. Este tipo de red es tolerante al ruido en los datos, porque Hopfield logró demostrar que el entrenamiento de su red se puede entender en términos de la energía del sistema. De tal forma, la red itera cuando hay ruido, buscando alcanzar el estado con energía mínima.

Hopfield jugó un papel fundamental en el resurgimiento del interés por las redes neuronales, pues no solo mostró la forma en la que conceptos de mecánica estadística podían aplicarse a las redes neuronales, sino que además, de alguna forma, legitimó la investigación en torno a las redes neuronales. Su trabajo daría un nuevo impulso a esta disciplina que regresaría para quedarse. En 2024, se le otorgó el Premio Nobel de física. El comité del Nobel indicó que Hopfield "inventó en 1982 una red que usa un método para almacenar y recrear patrones. Para ello, se inspiró en modelos de la física sobre la forma en la que muchas partes pequeñas en un sistema afectan al sistema como un todo. Este invento se volvió muy importante, por ejemplo, para el análisis de imágenes."

FIGURA 12: David E. Rumelhart.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

En 1982, David E. Rumelhart (ver figura 12) desarrolló el algoritmo de propagación hacia

atrás de forma independiente (recordemos que ya se habían hecho otras propuestas previas, las cuales Rumelhart desconocía), y se lo enseñó a Terrence Sejnowski quien lo probó y se percató de que era más rápido que las máquinas de Boltzmann que se usaban en ese entonces. Geoffrey Hinton (quien realizaba en ese entonces una estancia posdoctoral con Rumerhart en la Universidad de California en San Diego), fue uno de los primeros usuarios del algoritmo de propagación hacia atrás de Rumelhart, aunque en un principio estuvo un tanto reacio a abandonar el uso de las máquinas de Boltzmann. Sin embargo, después de un tiempo, Hinton se convenció de que el algoritmo de propagación hacia atrás tenía un mejor desempeño que las máquinas de Boltzmann.

La publicación de este algoritmo (Rumelhart *et al*., 1986) fue el segundo acontecimiento clave para el resurgimiento del interés en las redes neuronales.

La publicación del libro *Parallel Distributed Processing* (Rumelhart y McClelland, 1987), editado por David Rumelhart y James McClelland fue otro acontecimiento muy relevante para la comunidad de redes neuronales. En los dos volúmenes que conforman este libro, se usaba el término "conexionismo" para denominar a las redes neuronales y otros modelos similares de aprendizaje. El libro originó un fuerte debate entre la comunidad de IA simbólica y los "conexionistas". Se dice que Geoffrey Hinton llegó a decir que los símbolos era el "éter luminoso de la IA" (Russell y Norvig, 2021). En otras palabras, Hinton decía que los símbolos son un modelo de la inteligencia que no funciona y que nos lleva por la dirección equivocada.

En 1983, Geoffrey Hinton inventó (junto con Terrence Sejnowski) las máquinas de Boltzmann, las cuales pueden aprender a reconocer elementos característicos en un conjunto de datos. Hinton y Sejnowski aplicaron por primera vez las máquinas de Boltzmann a problemas de ciencia cognitiva (Hinton y Sejnowski, 1983). En 2024, se le otorgó a Hinton el Premio Nobel de Física. El comité del Nobel indicó que Hinton "usó herramientas de física estadística para crear la máquina de Boltzmann y que este invento se volvió muy importante para, por ejemplo, clasificar y crear imágenes". Otras contribuciones de Hinton a las redes neuronales incluyen: las representaciones distribuidas, la red neuronal con retardo y las máquinas de Holmholtz. Debido a sus muchas contribuciones a lo largo de más de 40 años, muchos lo consideran el "abuelo" del aprendizaje profundo. En años recientes se han vuelto famosas sus advertencias sobre los peligros de la IA.

En 1987 se crea un programa llamado NETtalk, el cual fue desarrollado para aprender a pronunciar texto escrito en inglés usando una red neuronal (Sejnowski y Rosenberg, 1987). NETtalk fue el resultado del trabajo de investigación realizado

por Terrence Sejnowski y Charles Rosenberg a mediados de la década de 1980. El objetivo de esta investigación era construir modelos simplificados que pudieran proporcionar algún tipo de evidencia sobre la complejidad de las tareas cognitivas involucradas en el aprendizaje humano. Se buscaba usar un modelo conexionista que fuese capaz de aprender a realizar una tarea comparable a la del humano. La red neuronal utilizada tenía tres capas y 18,629 pesos ajustables, lo cual era bastante grande para la época. El conjunto de datos para el entrenamiento consistió de 20,000 palabras obtenidas acompañadas de su pronunciación correspondiente. Esta red neuronal fue entrenada de dos formas: usando primero una máquina de Boltzmann y después el algoritmo de propagación hacia atrás. El éxito de NETtalk llevó al desarrollo de modelos mucho más complejos de aprendizaje basados en redes neuronales.

VII. EL SEGUNDO INVIERNO DE LA IA (1987-1993)

Aunque ya vimos que el primer "invierno de la IA" ocurrió en la década de 1970, esta expresión realmente se acuñó en 1984, durante un debate llevado a cabo en la reunión anual de la *American Association of Artificial Intelligence*. Durante este evento, Roger Schank y Marvin Minsky (dos investigadores experimentados en IA que habían sufrido las consecuencias del primer invierno), advirtieron a un grupo de empresarios sobre el crecimiento desmesurado del entusiasmo en la IA que se había dado a principios de la década de 1980. Schank y Minsky advirtieron que después del entusiasmo vendría la decepción y describieron una reacción en cadena similar a un "invierno nuclear", que comenzaría con el pesimismo de la comunidad de IA, seguido por el pesimismo de la prensa, el cual vendría acompañado de severos recortes al financiamiento de la investigación en IA (Crevier, 1993). Esta predicción resultó

FIGURA 13: Una *LISP Machine*.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

correcta, pues el segundo "invierno de la IA" comenzó solo tres años después de este evento, en 1987.

El entusiasmo por la IA que imperaba a principios de la década de 1980 parecía estar justificado. El primer sistema experto comercial, llamado XCON, desarrollado en 1978 por la Universidad Carnegie-Mellon para *Digital Equipment Corporation* (DEC) tuvo un enorme éxito. Este programa era un asistente para procesar pedidos de computadoras VAX, el cual seleccionaba de forma automática los componentes del sistema a partir de los requerimientos de cada cliente. El sistema comenzó a usarse en 1980 y solo seis años después, DEC estimaba que les había producido ahorros de 40 millones de dólares.

Varias corporaciones siguieron el ejemplo de DEC y comenzaron a instalar sistemas expertos. Para 1985, estas empresas habían invertido más de 1,000 millones de dólares en IA. Alrededor de estas aplicaciones comenzaron a crearse empresas para darles soporte. Esto incluyó a varias empresas de software como *Teknowledge* e *Intellicorp*, así como a empresas de hardware como *Symbolics* y *LISP Machines Inc.*, las cuales construían computadoras especializadas llamadas *LISP Machines* (ver figura 13).

Estas computadoras usaban el lenguaje de programación LISP como lenguaje nativo (o sea, en hardware), con lo cual podían ejecutar mucho más rápido los programas escritos en este lenguaje de programación, que en ese entonces era muy utilizado por los desarrolladores de algoritmos de IA.

Pero para 1987, el mercado de las *LISP Machines* colapsó. La razón fue que surgieron empresas como *Sun Microsystems* que comenzaron a ofrecer hardware más poderoso. Pronto, surgieron empresas, como *Lucid*, que comenzaron a ofrecer ambientes de LISP optimizados para las estaciones de trabajo que vendía *Sun Microsystems*.

El desempeño de las estaciones de trabajo de *Sun Microsystems* comenzó a poner en aprietos a los desarrolladores de las *LISP Machines*. Pronto, otras empresas como Apple e IBM comenzaron a ofrecer también arquitecturas más simples y populares para ejecutar programas en LISP. Para 1987, estas computadoras eran ya más poderosas que las costosas *LISP Machines*.

Aunado a esto, comenzaron a aparecer motores basados en reglas para computadoras de escritorio, tales como CLIPS, con lo cual las empresas no tenían ya razones para pagar más por computadoras especializadas para ejecutar LISP. Fue así como esta industria con valor de 500 millones de dólares, fue reemplazada en solo un año (Crevier, 1993).

A principios de la década de 1990, las empresas se percataron de que sus sistemas expertos, pese a ser exitosos, estaban teniendo costos de mantenimiento

demasiado elevados. Además, estos sistemas expertos eran difíciles de actualizar y eran propensos a errores si las reglas no se codificaban cuidadosamente. Poco a poco, las empresas dedicadas a desarrollar sistemas expertos fueron desapareciendo o diversificándose para poder sobrevivir.

Para 1993, más de 300 empresas que desarrollaban soluciones basadas en IA se habían ido a la quiebra.

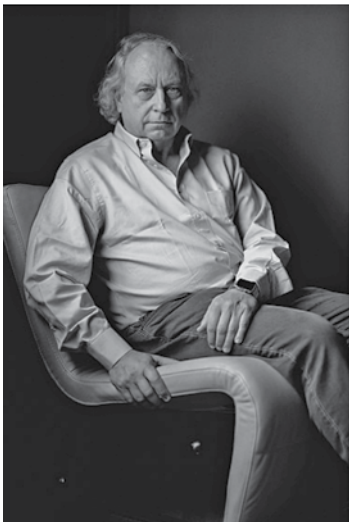
Un segundo evento que contribuyó al "segundo invierno de la IA" fue el fracaso del proyecto de la quinta generación impulsado por Japón. Después de invertir $850 millones de dólares a lo largo de diez años, el gobierno japonés se percató de que los resultados obtenidos estaban lejos de las ambiciosas metas que se trazaron originalmente.

En 1983, y en respuesta al proyecto de la quinta generación, DARPA comenzó a financiar la investigación en IA a través de la *Strategic Computing Initiative*. De origen, este programa fue diseñado con metas prácticas y alcanzables, las cuales incluso incluían a la inteligencia artificial general (es decir, el desarrollo de software de IA capaz de efectuar tareas distintas con la misma programación) como un objetivo a largo plazo. Este programa estuvo bajo la dirección de la *Information Processing Technology Office* (IPTO) y también se incluyó financiamiento para la investigación en supercómputo y en microelectrónica. Para 1985, ya se habían gastado $100 millones de dólares en 92 proyectos que se estaban llevando a cabo en 60 instituciones (la mitad de ellas en empresas y la otra mitad en universidades y laboratorios del gobierno de los Estados Unidos).

En 1987, Jack Schwarz fue colocado a la cabeza de IPTO y ahí comenzaron los problemas. Schwarz veía con malos ojos a los sistemas expertos y aplicó recortes profundos y brutales a los fondos dedicados a la IA. Schwarz estaba convencido de que la IA no era la apuesta más adecuada en investigación y pensaba que DARPA debía invertir en otras tecnologías que consideraba más prometedoras (McCorduck, 2004).

Otro fenómeno muy interesante que se produce durante el "segundo invierno de la IA" es la "rebelión" de varios científicos

FIGURA 14: Rodney Brooks.



FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

que comienzan a cuestionar fuertemente la "hipótesis de los símbolos físicos". La figura central de esta rebelión fue Rodney Brooks (ver figura 14), un investigador del Instituto Tecnológico de Massachusetts que publicó en 1990 un artículo en el cual cuestiona fuertemente lo que él denominó el "dogma" de la hipótesis de los símbolos físicos (Brooks, 1990). En este artículo Brooks argumenta que los símbolos no siempre son necesarios para producir IA, porque el mundo es nuestro mejor modelo. Estos cuestionamientos condujeron a enfoques distintos de desarrollar IA, sobre todo en áreas como robótica, los cuales resultaron muy exitosos.

También en 1990, Yann LeCun (ver figura 15), en Laboratorios Bell, usa una red neuronal convolucional para reconocer dígitos escritos a mano. Este sistema fue usado ampliamente en la década de 1990 para leer códigos postales y cheques personales. Muchos lo consideran como la primera aplicación práctica, genuinamente útil, de las redes neuronales (Russell y Norvig, 2021). Yann LeCun recibió, junto con Yoshua Bengio y Geoffrey Hinton el Premio Turing 2018, por sus contribuciones al aprendizaje profundo. Este es considerado como el premio más importante en computación.

FIGURA 15: Yann LeCun.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

VIII. LOS RESULTADOS PROMETIDOS (1993-2011)

En el período de 1993 a 2011, la IA finalmente comienza a cumplir algunas de sus viejas promesas, debido, sobre todo, a un mayor poder de cómputo, aunque también debido al desarrollo de algoritmos más poderosos.

Deep Blue fue el primer programa de computadora capaz de derrotar al campeón del mundo en ajedrez en 1997. *Deep Blue* derrotó a Garry Kasparov en dos partidas, perdió en una y empató en tres más, volviéndose un hito en la historia de la IA.

Después de este histórico triunfo, IBM decidió que *Deep Blue* no volvería a jugar jamás y la envió al Museo Smithsoniano en Washington, D.C.

El desarrollo de *Deep Blue* (ver figura 16) comenzó en 1985 en la Universidad Carnegie-Mellon y el proyecto se trasladó a IBM en 1989. El algoritmo de *Deep Blue* dependía de la fuerza bruta y se ejecutaba en una supercomputadora IBM RS/6000 SP que contaba con 32 procesadores diseñados para ejecutar un sistema experto para jugar ajedrez. El algoritmo de *Deep Blue* era capaz de evaluar 200 millones de posiciones de ajedrez por segundo.

FIGURA 16: Deep Blue.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

Durante la década de 1990, gana amplia aceptación un paradigma denominado "agentes inteligentes". Un agente inteligente es un sistema que percibe su ambiente y toma acciones que maximizan sus probabilidades de éxito.

Pero tal vez uno de los eventos más relevantes de este período haya sido la incorporación de múltiples disciplinas (p.ej., matemáticas, economía, investigación de operaciones e ingeniería eléctrica) a la IA.

Un libro publicado por Judea Pearl en 1988 incorpora la teoría de probabilidad en la IA (Pearl, 1988). Así, surgen diversas herramientas nuevas en la IA, tales como las redes Bayesianas, los modelos ocultos de Markov, la teoría de la información, el modelado estocástico y la optimización clásica.

En un artículo publicado en 1994, Lofti Zadeh (ver figura 17) de la Universidad de California en Berkeley, acuña el término "*soft computing*", en el cual incluye a la lógica difusa, las redes neuronales, los algoritmos evolutivos, el razonamiento probabilístico y la teoría del caos (Zadeh, 1994). En ese mismo año, James Bezdek (Siddique y Adeli, 2013) utiliza por primera vez el término "Inteligencia Computacional" (*Computational Intelligence*) para referirse a un sistema que lidia con datos de bajo nivel tales como los numéricos y que no usa una representación tradicional (o sea, de la IA simbólica) del conocimiento. Bezdek indicaría después que la Inteligencia Computacional se basa en el uso de técnicas de "*soft computing*", mientras que la IA tradicional usa técnicas de "*hard computing*". En 2003, el Instituto de Ingenieros Eléctricos y Electrónicos (IEEE por sus siglas en inglés) aprueba cambiar el nombre al *IEEE Neural Networks Council* por el de *IEEE Computational Intelligence Society*. Con este cambio, se indica que la Inteligencia Computacional incluye a la lógica difusa, las redes neuronales

y los algoritmos evolutivos, así como a los sistemas híbridos que combinan este tipo de técnicas. La Inteligencia Computacional jugaría un papel preponderante dentro de la IA durante el siglo XXI.

FIGURA 17: Lofti A. Zadeh.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

En 1997 ocurre otro evento de gran relevancia para el desarrollo del aprendizaje profundo: Sepp Hochreiter y Jürgen Schmidhuber proponen la denominada *Long Short-Term Memory* (LSTM). Éste es un tipo de red neuronal recurrente diseñada para mitigar un problema conocido como "el desvanecimiento del gradiente", que impedía entrenar redes neuronales profundas usando métodos de gradiente o la propagación hacia atrás (Hochreiter y Schmidhuber, 1997). Este mecanismo proporciona a las redes neuronales una memoria de corto plazo que puede durar miles de iteraciones (de ahí el nombre). La intuición detrás de la arquitectura de una LSTM es crear un modelo adicional en una red neuronal que sea capaz de aprender cuándo recordar y cuándo olvidar información pertinente. En otras palabras, la red aprende de forma efectiva qué información podría necesitar después en una secuencia, así como el momento en el que ya no necesita esa información. Schmidhuber es un personaje muy controversial, pues ha expresado públicamente que él y otros pioneros de la IA no han recibido el crédito debido por sus contribuciones al aprendizaje profundo y que han favorecido a Yoshua Bengio, Yann LeCun y Geoffrey Hinton.

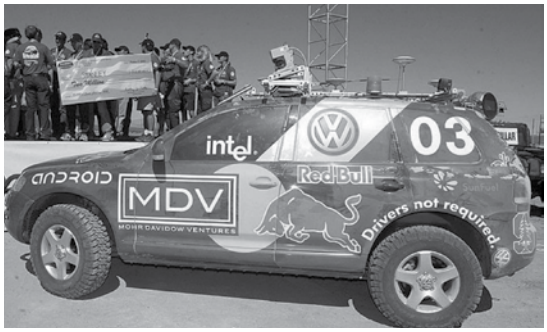
FIGURA 18: Jürgen Schmidhuber.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

En 2003, Yoshua Bengio publica un artículo donde utiliza redes neuronales para procesamiento en lenguaje natural (Bengio *et al.*, 2003). Entrenar una red neuronal para distinguir oraciones que tienen un significado de las que no tienen sentido era una tarea muy difícil porque existen muchas formas diferentes de expresar la misma idea usando lenguaje natural. Esto provoca la "maldición de la dimensionalidad" antes mencionada, lo cual demandaba

enormes cantidades de datos de entrenamiento. En esta publicación, Bengio y sus colegas introducen un nuevo tipo de representación (denominada *high-dimensional word embeddings*) para las palabras, la cual permitió un cambio de paradigma en el procesamiento en lenguaje natural y en la traducción automática. Posteriormente, Bengio propone mecanismos de "atención" que hacen que una red neuronal se enfoque únicamente en el contexto que sea relevante a una oración. Estos mecanismos permitirían una nueva generación de procesadores de lenguaje natural y de traductores mucho más poderosos.

FIGURA 19: Stanley.



FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

En 2004, DARPA decide establecer una competencia de vehículos autónomos con premios en efectivo. En su primera edición, denominada *DARPA Grand Challenge*, ningún vehículo pudo terminar la competencia. Pero en 2005, un equipo de la Universidad de Stanford logró ganar esta competencia, pues fueron capaces de conducir, de forma autónoma por el desierto, un vehículo llamado *Stanley* (ver figura 19) a lo largo de más 200 kilómetros.

En 2007, se establece el *DARPA Urban Challenge*, en el cual resulta ganador un equipo de la Universidad Carnegie-Mellon, al lograr que su vehículo navegara, de forma autónoma en un ambiente urbano, respetando las leyes de tránsito y exponiéndose a los riesgos del tráfico, a lo largo de casi 90 kilómetros.

Con esto, se cumplía otra de las viejas promesas de la IA.

IX. EL USO DE GRANDES CANTIDADES DE DATOS (2011-FECHA)

A inicios del siglo XXI, el acceso a grandes cantidades de datos y a un mucho mayor poder de cómputo, cambió la forma de desarrollar algoritmos de IA. Los avances en aprendizaje profundo (particularmente, en las redes neuronales convolucionales y en las redes neuronales recurrentes) permitieron grandes avances en el procesamiento de imágenes y video, en el análisis de textos y en el reconocimiento de voz.

La disponibilidad de grandes volúmenes de datos (a lo que se le conoce como *big data*) ha hecho necesario desarrollar nuevos modelos de procesamiento, a fin de extraer información útil de ellos. La disponibilidad de grandes volúmenes de datos permitió que las redes neuronales profundas pudiesen producir resultados que no terminan de sorprendernos (sobre todo en años recientes). A continuación mostraremos algunos ejemplos de ello.

ALEXNET

AlexNet es el nombre de una arquitectura de red neuronal convolucional diseñada por Alex Krizhevsky en colaboración con Ilya Sutskever y Geoffrey Hinton (quien fue el asesor de tesis de Krizhevsky). AlexNet compitió en el *ImageNet Large Scale Visual Recognition* Challenge el 30 de septiembre de 2012 y tuvo un error de solo 15.3%, lo cual la colocó a casi 11 puntos de distancia del segundo lugar.

AlexNet tenía 60 millones de parámetros y 650,000 neuronas. En el artículo en el que se explica AlexNet (Krizhevsky *et al.*, 2017), se indica que la profundidad del modelo fue esencial para lograr estos resultados. Esto implicó un alto costo computacional, para lo cual se usaron GPUs (*Graphics Processing Units*) durante el entrenamiento de la red.

AlexNet se considera un referente en cuanto al éxito del "aprendizaje profundo" en tareas de clasificación complejas.

LAS REDES GENERATIVAS ADVERSARIAS

En 2014, Yoshua Bengio (ver figura 20) desarrolló junto con Ian Goodfellow (quien en ese entonces era su estudiante de doctorado) el concepto de "redes generativas adversarias" (Goodfellow *et al.*, 2014). Mientras que la mayoría de las redes se diseñaron para reconocer patrones, una red generativa aprende a generar objetos que son difíciles de distinguir de entre los que están disponibles en el conjunto de entrenamiento. Se llama "adversaria" a esta técnica, porque es posible entrenar

FIGURA 20: Yoshua Bengio.

FUENTE: Wikipedia (disponible bajo una licencia de *Creative Commons*).

una de estas redes para generar salidas falsas creíbles contra otra red que aprenda a identificar salidas falsas. Esto permite un proceso de aprendizaje dinámico inspirado en la teoría de juegos. Este proceso se usa frecuentemente en el aprendizaje no supervisado. En años recientes, este tipo de redes se han usado, por ejemplo, para generar fotografías muy realistas que contienen a personas inexistentes.

ALPHAGO

AlphaGo es un programa diseñado por *DeepMind Technologies* para jugar Go, que es un juego muy antiguo con un espacio de búsqueda más grande que el del ajedrez.

Para aprender a jugar, AlphaGo tomó como entrada una descripción del tablero de Go y la procesó a través de una red neuronal profunda. Usando la denominada "red de políticas", selecciona el siguiente movimiento. Luego, otra red (llamada "red de valor") predice al ganador del juego. AlphaGo fue entrenado con un gran número de partidas entre personas a fin de que desarrollara un estilo de juego similar al de los humanos. Luego, jugó muchísimas veces contra sí mismo y mediante "aprendizaje por refuerzo", logró adquirir un nivel de juego muy elevado. En octubre de 2015, AlphaGo jugó su primera partida contra un humano: el tres veces Campeón Europeo, Fan Hui. AlphaGo derrotó a Hui en las cinco partidas que jugaron.

En 2016, AlphaGo se enfrentó al legendario jugador coreano Lee Sedol, ganador de 18 títulos mundiales y considerado como el más grande jugador de Go de esa década. AlphaGo ganó 4 partidas y perdió solo una en el torneo jugado en Seúl, Corea del Sur, en marzo de 2016. Este torneo fue visto por más de 200 millones de personas y marcó un nuevo hito en los juegos por computadora. A diferencia de Deep Blue, AlphaGo utiliza técnicas de aprendizaje máquina de propósito mucho más general y muchos consideran a este programa como un paso muy importante hacia la IA General.

Deepmind Technologies es una empresa inglesa que fue creada en 2010 por Demis Hassabis, Shane Legg y Mustafa Suleyman. En 2014, esta empresa fue adquirida por Google, que pagó 400 millones de libras esterlinas por ella. Hassabis permaneció como su Vicepresidente de Ingeniería.

El predecesor de AlphaGo fue AlphaZero que, en 2017, logró derrotar a StockFish 8 que, en ese entonces, era el campeón mundial de ajedrez por computadora (con un ELO de 3400). AlphaZero solo requirió ser entrenado durante 4 horas jugando contra sí mismo.

En 2020, Demis Hassabis y John Jumper presentaron AlphaFold2. Esta red neuronal pudo predecir la estructura de virtualmente todas las proteínas conocidas (200 millones). Esto ha permitido a los científicos entender mejor cuestiones como la resistencia a los antibióticos y les ha permitido crear enzimas que descomponen el plástico.

Este trabajo los haría ganar (junto con David Baker), el Premio Nobel de Química en 2024.

CHATGPT

ChatGPT es un chatbot de inteligencia artificial generativa diseñado por *OpenAI* que pertenece a la familia de modelos de lenguaje conocidos como "*Generative Pre-trained Transformer*" (GPT). ChatGPT se basa en un *Large Language Model* (LLM) llamado GPT-4o y puede generar respuestas y sostener conversaciones similares a las de un humano.

ChatGPT fue entrenado por medio de humanos y de aprendizaje por refuerzo usando la infraestructura de supercómputo llamada Azure, de Microsoft. Se liberó al público en general en noviembre de 2022, causando enorme conmoción en todo el mundo. La versión 4.0 fue liberada el 14 de marzo de 2023 para los usuarios de ChatGPT plus (la versión comercial del programa).

Se dice que ChatGPT catapultó el interés en la IA, atrayendo más inversiones y un gran interés del público en general en esta disciplina. Pero también ha despertado preocupaciones pues su uso inadecuado puede dar lugar a desinformación y a plagio, entre otras muchas cosas.

OpenAI se fundó en 2015 como una organización sin fines de lucro, con la misión de asegurar que la inteligencia artificial general beneficiara a toda la humanidad. Entre sus fundadores se encontraban: Elon Musk, Sam Altman, Ilya Sutskever, Greg Brockman y otros expertos en IA. Entre sus inversionistas se encuentra Microsoft que, además de aportar miles de millones de dólares, le ha dado acceso a la empresa a su plataforma de supercómputo (Azure).

EL SUEÑO DE LA INTELIGENCIA ARTIFICIAL GENERAL

Los avances en aprendizaje máquina han hecho resurgir el interés en la denominada "inteligencia artificial general", que se define como la habilidad de resolver cualquier problema con un algoritmo de IA, en vez de limitarse a encontrar la solución a un problema en particular.

Los modelos fundacionales, que son modelos de IA de gran escala, entrenados con grandes cantidades de datos no etiquetados que pueden usarse para una gran cantidad de tareas, comenzaron a desarrollarse en 2018. Modelos tales como **GPT-3**, liberado por *OpenAI* en 2020 y **Gato**, liberado por *DeepMind* en 2022, han sido descritos como contribuciones seminales a la inteligencia artificial general.

En 2023, *Microsoft Research* probó **GPT-4** en una gran variedad de tareas y concluyó que este sistema "podía ser visto, de forma razonable, como una primera versión (todavía incompleta) de un sistema de IA general."

Gato es una red neuronal profunda multimodal desarrollada por *DeepMind* que usa la misma red con los mismos pesos para jugar Atari, para etiquetar imágenes, para apilar bloques con un brazo de robot, para dialogar con un humano (chats) y para alrededor de 600 tareas, tomando decisiones con base en su contexto. Su red neuronal utiliza 1,200 millones de parámetros.

Flamingo es un modelo de lenguaje con apoyo visual desarrollado por *DeepMind* y cuenta con 80,000 millones de parámetros. Es capaz de tomar una imagen o video como entrada y, con base en ésta, tener una conversación con un humano (mediante un chat) y responder preguntas no triviales sobre la imagen o el video de manera muy acertada.

X. EL FUTURO DE LA INTELIGENCIA ARTIFICIAL

Evidentemente, la creciente popularidad de la que goza la IA actualmente, ha vuelto muy optimistas a los especialistas del área, quienes vislumbran grandes cambios en los próximos años en varias áreas, de entre las que destacan las siguientes:

- Educación: Los estudiantes podrían recibir contenidos diseñados específicamente para sus necesidades particulares. La IA también podría determinar las estrategias educativas óptimas para cada estudiante. Algunos creen que para 2028, el sistema educativo será radicalmente distinto del actual.
- Salud: Algunos piensan que será posible contar con tratamientos personalizados (basados en nuestro ADN) en el futuro. Hoy en día, existen ya terapias personalizadas producidas gracias a la IA, pero se esperan avances mucho más radicales en los próximos años.
- Finanzas: El procesamiento en lenguaje natural combinado con el aprendizaje de máquina podría permitir a los bancos y a los asesores financieros el

poder ayudar de forma mucho más eficiente a sus clientes en diversas áreas: detección de fraudes, planeación financiera, servicio al cliente, adquisición de seguros y monitoreo de puntaje de crédito, entre muchas otras.

- **Nuevas formas de hacer investigación científica:** Se espera lograr avances científicos muy significativos en los próximos años con la ayuda de la IA debido a su capacidad de analizar grandes cantidades de datos y descubrir patrones y relaciones complejas en ellos.
- **Transporte:** Se cree que los vehículos autónomos se volverán populares tanto para transporte comercial como para transporte público, produciendo un cambio radical en la forma de transportar mercancías y personas.

Pero también hay quienes avizoran consecuencias negativas del uso de la IA en el futuro. Por ejemplo:

- **El fin de la privacidad:** Los sistemas que usan IA requieren grandes cantidades de información privada para poder operar. De tal forma, el desarrollo de aplicaciones más sofisticadas que nos faciliten nuestra vida diaria requerirá de una constante invasión a nuestra privacidad, lo cual ha iniciado ya fuertes debates en algunos países.
- **Aplicaciones militares:** ¿Se desarrollarán sistemas de misiles que sean controlados por IA sin intervención humana? ¿Cuáles son los límites éticos de la IA?
- **Pérdida de empleos:** Muchos temen sobre la potencial pérdida de empleos que acarreará la IA. Los expertos dicen que se perderán empleos tediosos y carentes de creatividad, pero no resulta claro todavía cuál será el impacto real de la IA en la economía de cada país.
- **Impacto ambiental:** Una de las preocupaciones que ha surgido en años recientes es el impacto ambiental de las aplicaciones de IA que hacen uso de enormes recursos de cómputo. Un estudio reciente de investigadores de la Universidad de California en Riverside reveló, por ejemplo, que Microsoft utilizó aproximadamente 700,000 litros de agua durante el entrenamiento de GPT-3. Esa cantidad de agua es equivalente a la requerida para producir 370 autos BMW o 320 vehículos Tesla. La razón para este consumo es que se utilizan enormes cantidades de energía y al convertirse en calor, se requieren grandes cantidades de agua para mantener las temperaturas de los equipos de cómputo bajo control. ChatGPT también consume cantidades importantes de agua cuando se le hacen consultas o cuando se le pide que genere texto. Para una simple conversación de unas 20 a 50 preguntas, el agua consumida equivale a una botella de 500 ml. Esta cantidad se vuelve relevante si consideramos que ChatGPT tiene millones de usuarios en el mundo.

XI. EL MIEDO A LA INTELIGENCIA ARTIFICIAL

Prácticamente desde sus orígenes, la IA ha generado miedo no solo entre el público en general, sino también entre expertos (uno de los ejemplos más notables es Geoffrey Hinton, quien ha expresado públicamente su temor a un uso inadecuado de la IA).

El mayor temor es que los algoritmos basados en IA se salgan de control y acaben por exterminar a la humanidad. La mayoría de estos temores son infundados, pero claramente siempre hay riesgos asociados al uso de nuevas tecnologías y la IA no es la excepción. Por ello, el tipo de tareas en las que se usa la IA es un elemento clave para reducir riesgos, pues es claro que, al igual que cualquier otra tecnología avanzada, la IA puede ser peligrosa si se le utiliza para ciertas tareas (p.ej., para desarrollar armas biológicas).

Otro elemento importante a tomarse en cuenta es el hecho de que las computadoras no son seres vivos y tampoco tienen conciencia. Eso permite a un humano borrar un programa peligroso o apagar una computadora que pudiera causar daño a alguna persona. La mayoría de los temores actuales se deben a la ciencia ficción, que suele exagerar los avances tecnológicos, haciendo ver más peligrosas de lo que realmente son a las tecnologías actuales y a las que surgirán en los próximos años.

Sin embargo, es importante contar con un marco legal adecuado que permita regular el uso de nuestros datos personales, a fin de reducir el riesgo de que éstos se usen de manera inadecuada en programas de IA. También es importante introducir cursos de ética desde el nivel básico, pues las nuevas generaciones crecerán con esta tecnología y deben tener una clara concepción de los riesgos involucrados en su uso inadecuado o inmoral.

XII. CONCLUSIÓN

En general, vivimos en una era en la que contamos con programas muy poderosos que utilizan grandes cantidades de datos y un enorme poder de cómputo para realizar tareas usando algoritmos de IA, con resultados impresionantes. Se cree que en el futuro, la IA podrá ayudarnos a resolver problemas fundamentales para nuestra supervivencia tales como: la falta de agua, el cambio climático y la sequía, entre muchos otros.

Claramente, la IA subsimbólica (que usa números para almacenar el conocimiento y depende de grandes cantidades de datos) ha triunfado sobre la

IA simbólica de la década de 1950. Sin embargo, uno de los grandes retos es el de poder explicar las soluciones que estos sofisticados algoritmos producen. Y ahí, la IA simbólica puede ser de mucha ayuda.

Y, evidentemente, seguirá (y seguramente aumentará) el temor a las consecuencias del uso de estos algoritmos de IA, sobre todo si éstos se emplean en tareas críticas.

XIII. BIBLIOGRAFÍA

[Amari, 1967] Amari, Shunichi (June 1967). "A Theory of Adaptive Pattern Classifiers", *IEEE Transactions on Electronic Computers*, EC-16(3): 299-307.

[Anderson y Rosenfeld, 2000] Anderson, James A.; Rosenfeld, Edward, Eds. (2000). *Talking Nets: An Oral History of Neural Networks*, The MIT Press, USA, ISBN 978-026251111-7.

[Bengio *et al.*, 2003] Bengio, Yoshua; Ducharme, Réjean; Vincent, Pascal ; Janvin, Christian (March 2003). "A Neural Probabilistic Language Model", *The Journal of Machine Learning Research*. 3:1137-1155.

[Brooks, 1990] Brooks, Rodney A. (1990). "Elephants don't play chess", *Robotics and Autonomous Systems*, 6(1-2): 3-15.

[Buchanan y Feigenbaum, 1978] Buchanan, B.G.; Feigenbaum, E.A. (1978). "DENDRAL and Meta-DENDRAL: Their Applications Dimension", *Artificial Intelligence*, 11: 5-24.

[Crevier, 1993] Crevier, Daniel (1993). *AI: The Tumultuous Search for Artificial Intelligence*, Basic Books, New York, USA. ISBN 978-0-465-02997-6.

[Feigenbaum y Shrobe, 1993] Feigenbaum, Edward; Shrobe, Howard (1993). "The Japanese Fifth Generation Project: Introduction, survey, and evaluation", *Future Generation Computer Systems*, 9: 105-117.

[Fogel *et al.*, 1966] Fogel, L.J.; Owens, A.J. ; Walsh, M.J. (1966). *Artificial Intelligence Through Simulated Evolution*, Wiley, New York, USA.

[Gelernter, 1959] H. Gelernter (1959). "Realization of a Geometry-Theorem Proving Machine". In *Proceedings of an International Conference on Information Processing*, pp. 273-282. UNESCO House.

[Goodfellow *et al.*, 2014] Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). "Generative Adversarial Nets", In *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*, pp. 2672-2680.

[Hebb, 1949] Hebb, Donald (1949). *The Organization of Behavior. A Neuropsychological Theory*, Wiley, New York, USA, ISBN 978-1-135-63190.

[Hinton y Sejnowski, 1983] Hinton, Geoffrey E.; Sejnowski, Terrence J. (June 1983). "Optimal Perceptual Inference", In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Washington, D.C. USA, IEEE Computer Society, pp. 448-453.

[Hochreiter y Schmidhuber, 1997] Hochreiter, Sepp; Schmidhuber, Jürgen (1997). "Long short-term memory", *Neural Computation*, 9(8): 1735-1780.

[Holland, 1962] Holland, J.H. "Outline for a logical theory of adaptive systems", *Journal of the Association for Computing Machinery*, 3: 297-314.

[Hopfield, 1982] Hopfield, J.J. (April 15, 1982). "Neural networks and physical systems with emergent collective computational abilities", *Proceedings of the National Academy of Sciences*, 79(8): 2554-2558.

[Ivakhnenko y Lapa, 1967] Ivakhnenko, A.G.; Lapa, V.G. *Cybernetics and Forecasting*. Elsevier, New York, USA.

[Kacprzyk y Pedrycz, 2015] Kacprzyk, Janusz; Pedrycz, Witold, Eds. (2015). *Springer Handbook of Computational Intelligence*, Springer, ISBN 978-3662435045.

[Kirdin *et al*., 2022] Kirdin, Alexander; Sidorov, Sergey; Zolotykh, Nikolai (2022). "Rosenblatt's First Theorem and Frugality of Deep Learning", *Entropy*, 24(11): 1635.

[Krizhevsky *et al*., 2017] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. (24 May 2017). "ImageNet Classification with Deep Convolutional Neural Networks", *Communications of the ACM*. 60(6): 84-90.

[Linnainmaa, 1970] Linnainmaa, Seppo (1970). *The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors*, Tesis de Maestría, University of Helsinki (en Finlandés).

[Linnainmaa, 1976] Linnainmaa, Seppo (1976). "Taylor expansion of the accumulated rounding error", *BIT Numerical Mathematics*, 16(2): 146-160.

[McCulloch y Pitts, 1943] McCulloch, Warren S.; Pitts, Walter (1943). "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, 5(4). 115-133.

[Mata, 2006] Mata Alonso-Lasheras, Santiago (2006). *El hombre que demostró el cristianismo: Ramon Llull*. Ediciones Rialp, España, ISBN 978-84-321-6573-3.

[McCarthy, 1958] McCarthy, John (1958). "Programs with Common Sense". In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, pp. 75-91, Her Majesty's Stationary Office, London, UK.

[McCorduck, 2004] McCorduck, Pamela (2004). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. Second Edition, A.K. Peters, ISBN 978-1-56881-205-2.

[Newell y Simon, 1976] Newell, Allen; Simon, Herbert A. (1976). "Computer science as empirical inquiry: symbols and search", *Communications of the ACM*, 19(3): 113-126.

[Pearl, 1988] Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, California, USA. ISBN 978-0-08-051489-5.

[Rechenberg, 1965] Rechenberg, I. (1965). "Cybernetic solution path of an experimental problem", Technical Report Library 34 Translation 1122, Royal Air Force Establishment.

[Robinson, 1965] Robinson, John Alan (1965). "A Machine-Oriented Logic Based on the Resolution Principle", *Journal of the ACM*, 12(1): 23-41.

[Rosenblatt, 1958] Rosenblatt, Frank (1958). "The perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, 65(6): 386-408.

[Rosenblatt, 1962] Rosenblatt, Frank (1962). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington, D.C., USA.

[Rumelhart *et al*., 1986] Rumelhart, David E. ; Hinton, Geoffrey E. ; Williams, Ronald J. (October 1986). "Learning representations by back-propagating errors". *Nature*. *323*(6088): 533-536.

[Rumelhart y McClelland, 1987] Rumelhart, David E.; McClelland, James L. (Editors). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition: Foundations*. MIT Press, USA, ISBN 978-0-26-268053-0.

[Russell y Norvig, 2021] Stuart J. Russell; Peter Norvig (2021). *Artificial Intelligence. A Modern Approach*, Fourth Edition, Pearson, USA, ISBN 978-0-13-461099-3.

[Samuel, 1959] Samuel, Arthur (1959). "Some Studies in Machine Learning the Game of Checkers", *IBM Journal of Research and Development*. 3(3):210-229.

[Schwefel, 1965] Schwefel, H.-P. (1965). *Kybernetische Evolution als Strategie der experimentellen Forschung* 36 *in der Strömungstechnik*, Dipl.-Ing. Thesis, Technical University ofBerlin, Hermann Föttinger--37 Institute for Hydrodynamics.

[Sejnowski y Rosenberg, 1987] Sejnowski, Terrence J. ; Rosenberg, Charles R. (1987). "Parallel networks that learn to pronounce English text", *Complex Systems*, 1(1):145-168.

[Shannon, 1948] Shannon, Claude E. (Julio 1948). "A Mathematical Theory of Communications", *The Bell System Technical Journal*, 27(3): 379-423.

[Siddique y Adeli, 2013] Siddique, Nazmul; Adeli, Hojjat (2013). *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*. John Wiley & Sons. ISBN 978-1-118-53481-6.

[Slagle, 1961] Slagle, James (1961). A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus, Symbolic Automatic INTegrator (SAINT), PhD thesis, Massachusetts Institute of Technology, June 1961.

[Sutton y Barto, 2018] Sutton, Richard S.; Barto, Andrew G. (1990). *Reinforcement Learning: An Introduction*. Second Edition MIT Press, Cambridge, Massachusetts, USA. ISBN 978-0-262-03924-6.

[Turing, 1936] Turing, A.M. (1936). "On computable numbers, with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, 42:230-265.

[Turing, 1950] Turing A.M. (1950). "Computing machinery and intelligence", *Mind*, 59: 433-460.

[Weizenbaum, 1966] Weizenbaum, Joseph (January 1966). "ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine", Communications of the ACM, 9(1): 36-45.

[Weizenbaum, 1976] Weizenbaum, Joseph (1976). *Computer Power and Human Reason: From Judgement to Calculation*, W.H. Freeman & Co., USA, ISBN 978-0716704638.

[Werbos, 1982] Werbos, Paul (1982). "Applications of Advances in Nonlinear Sensitivity Analysis", In R.F. Drenick and F. Kozin, Eds., *System Modeling and Optimization. Proceedings of the 10th IFIP Conference, New York City, USA, USA, August 31-September 4, 1981*, Springer-Verlag, Lecture Notes in Control and Information Sciences.

[Whitehead y Russell, 1910] Whitehead, Alfred North; Russell, Bertrand (1910). *Principia Mathematica*, Vol. 1, Cambridge University Press, Cambridge, UK.

[Widrow y Hoff, 1960] Widrow, B.; Hoff, M.E. (1960). "Adaptive switching circuits", in *1960 IRE WECON Convention Record*, pp. 96-104.

[Wiener, 1948] Wiener, Norbert (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, Massachusetts, Estados Unidos, ISBN 978-0-262-73009-9.

[Winograd y Cowan, 1963] Winograd, S.; Cowan, J.D. (1963). *Reliable Computation in the Presence of Noise*, MIT Press, USA.

[Zadeh, 1965] Zadeh, L.A. (June 1965). "Fuzzy Sets", *Information and Control*, 8(3): 338-353.

[Zadeh, 1994] Zadeh, Lofti A. (1 March 1994). "Fuzzy Logic, Neural Networks, and Soft Computing", *Communications of the ACM*, 37(3): 77-84.

LA ERA DE LA INTELIGENCIA ARTIFICIAL Y LOS DESAFÍOS EN AMÉRICA LATINA

SUMARIO

I. Introducción a la inteligencia artificial **II.** El panorama de la IA en diferentes regiones del mundo **III.** Desafíos tecnológicos **IV.** Desafíos sociales y culturales **V.** Consideraciones finales **VI.** Anexo: Apéndices **VII.** Bibliografía

> AMIKAM YALOVETZKY

Senior Manager – Corporate Marketing Communications en MediaTek. Ingeniero en Electrónica y Comunicaciones de la Escuela de Ingeniería de Jerusalén en Israel, es un capitán retirado de la Fuerza Aérea de Israel y ha recibido dos premios como Autor más vendido internacional de Amazon.

> HUGO SIMG ATILANO

Latin America Business Development Director – MediaTek. Ingeniero en Electrónica y Comunicaciones por el ITESM Campus Estado de México, cuenta con más de 33 años de experiencia. Fue profesor de ingeniería en el ITESM CCM y en la industria de Telecomunicaciones ha desempeñado actividades comerciales y de desarrollo de negocios, con historial de éxito demostrado colaborando en empresas multinacionales fabricantes y desarrolladores de equipos de medición electrónica, dispositivos y sistemas de comunicación móvil, y actualmente en Sistemas en Chip semiconductores.

I. INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

HISTORIA Y EVOLUCIÓN DE LA IA

La inteligencia artificial (IA) ha recorrido un largo camino desde sus inicios en la década de 1950. Inicialmente, la IA se centró en resolver problemas específicos y realizar tareas que requerían inteligencia humana, como el ajedrez. Hoy en día, la IA se ha expandido a diversas aplicaciones, desde asistentes virtuales hasta vehículos autónomos. Existen varios tipos principales de IA y las más conocidas son la IA Predictiva y la IA Generativa. La IA se basa en algoritmos y modelos matemáticos que permiten a las máquinas aprender de los datos y tomar decisiones o presentar datos a tomadores de decisiones. Con avances en el aprendizaje automático y el procesamiento del lenguaje natural, la IA está transformando industrias y mejorando la eficiencia en múltiples sectores. Sin embargo, también plantea desafíos éticos y de seguridad que deben ser abordados.

PRINCIPIOS Y TÉRMINOS BÁSICOS DE LA IA

La IA se basa en varios componentes fundamentales que permiten a las máquinas aprender y tomar decisiones:

- **Algoritmos:** Los algoritmos son conjuntos de reglas y procedimientos que las máquinas siguen para resolver problemas y realizar tareas. En la IA, los algoritmos de aprendizaje automático (machine learning) y aprendizaje profundo (deep learning) son especialmente importantes.

- Datos: Los datos son el combustible de la IA. Los sistemas de IA aprenden a partir de grandes volúmenes de datos, que pueden incluir texto, imágenes, audio y video. La calidad y cantidad de los datos son cruciales para el rendimiento de los modelos de IA.
- Modelos: Los modelos son representaciones matemáticas de los problemas que la IA intenta resolver. Los modelos de aprendizaje automático se entrenan utilizando datos para hacer predicciones o tomar decisiones. Los modelos de aprendizaje profundo, como las redes neuronales, son más complejos y pueden manejar grandes cantidades de datos no estructurados.

TÉCNICAS DE IA

Existen varias técnicas y enfoques en la IA, cada uno con sus propias fortalezas y aplicaciones:

- Aprendizaje Supervisado: En el aprendizaje supervisado, los modelos se entrenan utilizando datos etiquetados, donde cada entrada tiene una salida correspondiente. El objetivo es aprender una función que mapee las entradas a las salidas correctas.
- Ejemplos incluyen la clasificación de correos electrónicos como spam o no spam y el reconocimiento de imágenes.
- Aprendizaje No Supervisado: En el aprendizaje no supervisado, los modelos se entrenan utilizando datos no etiquetados. El objetivo es encontrar patrones o estructuras ocultas en los datos. Ejemplos incluyen la agrupación de clientes en segmentos de mercado y la detección de anomalías en datos financieros.
- Aprendizaje por Refuerzo: En el aprendizaje por refuerzo, los modelos aprenden a tomar decisiones mediante la interacción con un entorno. El objetivo es maximizar una recompensa acumulada a lo largo del tiempo. Ejemplos incluyen el entrenamiento de robots para realizar tareas específicas y el desarrollo de agentes de IA que juegan videojuegos

TIPOS DE IA

IA Generativa

La inteligencia artificial generativa es una subdisciplina de la IA que se centra en la creación de contenido nuevo y original a partir de datos existentes. Utilizando modelos avanzados como las redes generativas adversariales (GAN) y los transformadores, la IA generativa puede producir imágenes, texto, música y otros tipos de contenido que son indistinguibles de los creados por humanos.

Las GAN, por ejemplo, consisten en dos redes neuronales que compiten entre sí: una generadora que crea contenido y una discriminadora que evalúa su autenticidad. Este proceso iterativo mejora continuamente la calidad del contenido generado. Los modelos de transformadores, como GPT-3, utilizan grandes volúmenes de datos textuales para generar respuestas coherentes y contextualmente relevantes en lenguaje natural. La IA generativa tiene aplicaciones en una amplia gama de campos, desde el arte y el entretenimiento hasta la medicina y la ingeniería, permitiendo la creación de prototipos rápidos, la personalización de experiencias y la innovación en el diseño de productos. Sin embargo, también plantea desafíos *éticos* y de seguridad, como la generación de desinformación y la necesidad de garantizar el uso responsable de estas tecnologías.

IA Predictiva

La inteligencia artificial predictiva es una rama de la IA que se enfoca en el análisis de datos históricos para hacer predicciones sobre eventos futuros. Utilizando técnicas de aprendizaje automático y modelos estadísticos, la IA predictiva puede identificar patrones y tendencias en grandes conjuntos de datos, permitiendo a las organizaciones anticipar comportamientos y tomar decisiones informadas. Esta tecnología es ampliamente utilizada en diversos sectores, como la salud, donde puede predecir brotes de enfermedades y personalizar tratamientos; en las finanzas, para la detección de fraudes y la gestión de riesgos; y en el comercio, para prever la demanda de productos y optimizar inventarios. Los modelos predictivos, como los árboles de decisión, las redes neuronales y los modelos de regresión, se entrenan con datos históricos y se validan con datos nuevos para asegurar su precisión y fiabilidad. La IA predictiva no solo mejora la eficiencia operativa y reduce costos, sino que también proporciona una ventaja competitiva al permitir a las empresas y organizaciones responder proactivamente a las oportunidades y desafíos futuros.

Otros tipos de IA y términos de la industria

Existen varios tipos de inteligencia artificial, cada uno con diferentes niveles de complejidad y capacidades. La "IA débil" o "IA estrecha" se especializa en realizar tareas específicas y limitadas, como los asistentes virtuales (Siri, Alexa) y los sistemas de recomendación en plataformas de streaming. Esta forma de IA no tiene conciencia ni comprensión general del mundo. Por otro lado, la "IA fuerte" o "IA general" aspira a replicar la inteligencia humana en su totalidad, con la capacidad de entender, aprender y aplicar conocimientos en una amplia

variedad de tareas; sin embargo, esta forma de IA aún es teórica y no se ha logrado en la práctica. Además, se habla de "IA reactiva", que responde a estímulos específicos sin almacenar experiencias pasadas, y "IA con memoria limitada", que puede utilizar experiencias pasadas para tomar decisiones futuras. Finalmente, la "IA de teoría de la mente" y la "IA autoconsciente" representan niveles más avanzados y especulativos, donde las máquinas no solo entenderían las emociones y pensamientos humanos, sino que también tendrían una forma de autoconciencia. Cada tipo de IA tiene aplicaciones y desafíos únicos, y su desarrollo continuara transformando múltiples aspectos de la industria.

- IA agéntica, este tipo de IA representa una evolución significativa en el campo de la inteligencia artificial, donde los sistemas no solo responden a comandos predefinidos, sino que también actúan de manera autónoma para alcanzar objetivos específicos. Estos agentes inteligentes son capaces de percibir su entorno, tomar decisiones informadas y aprender de sus experiencias para mejorar su desempeño con el tiempo. A diferencia de las IA tradicionales, que requieren una programación explícita para cada tarea, la IA agentica utiliza algoritmos avanzados de aprendizaje automático y redes neuronales para adaptarse a situaciones nuevas y complejas. Esta capacidad de auto optimización y adaptación permite a los agentes agenticos desempeñar roles cruciales en una variedad de aplicaciones, desde la gestión de recursos en ciudades inteligentes hasta la personalización de servicios en plataformas digitales. Al integrar la IA agentica en nuestras vidas, estamos no solo mejorando la eficiencia y la efectividad de los sistemas, sino también abriendo nuevas posibilidades para la innovación y el progreso en múltiples sectores.
- Machine Learning: El aprendizaje automático, o machine learning, es una subdisciplina de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las máquinas aprender y mejorar a partir de la experiencia sin ser explícitamente programadas para cada tarea. Utilizando grandes volúmenes de datos, estos algoritmos identifican patrones y relaciones ocultas, lo que les permite hacer predicciones o tomar decisiones basadas en nuevos datos. Existen varios tipos de aprendizaje automático, incluyendo el aprendizaje supervisado, donde los modelos se entrenan con datos etiquetados; el aprendizaje no supervisado, que busca descubrir estructuras ocultas en datos no etiquetados; y el aprendizaje por refuerzo, donde los agentes aprenden a tomar decisiones mediante la interacción con su entorno y la maximización de recompensas acumuladas. El machine learning tiene aplicaciones en una amplia gama de campos, desde

la detección de fraudes y el diagnóstico médico hasta la personalización de contenido y la conducción autónoma, transformando industrias y mejorando la eficiencia y precisión de numerosos procesos.

- Deep Learning: El aprendizaje profundo, o deep learning, es una rama avanzada del aprendizaje automático que utiliza redes neuronales artificiales con múltiples capas para modelar y entender datos complejos. Inspiradas en la estructura y funcionamiento del cerebro humano, estas redes neuronales profundas son capaces de aprender representaciones jerárquicas de los datos, lo que les permite identificar patrones y características a diferentes niveles de abstracción. A diferencia de los métodos tradicionales de aprendizaje automático, que a menudo requieren la extracción manual de características, el aprendizaje profundo puede automatizar este proceso, lo que lo hace especialmente eficaz para tareas como el reconocimiento de imágenes, el procesamiento del lenguaje natural y la traducción automática. Gracias a los avances en hardware, como las unidades de procesamiento gráfico (GPU), y a la disponibilidad de grandes conjuntos de datos, el aprendizaje profundo ha logrado resultados impresionantes en diversas aplicaciones, superando en muchos casos el rendimiento humano. Esta tecnología está revolucionando campos como la visión por computadora, la robótica, la medicina y la inteligencia artificial general, abriendo nuevas posibilidades para la innovación y el desarrollo tecnológico.

IA EN LA NUBE, EN EL EDGE E HIBRIDA

IA en la Nube

La inteligencia artificial en la nube está transformando la manera en que las empresas y organizaciones implementan y escalan soluciones de IA, ofreciendo acceso a potentes recursos computacionales y herramientas avanzadas sin la necesidad de infraestructura local costosa. Aplicaciones como el análisis de grandes volúmenes de datos, la automatización de procesos empresariales, la personalización de experiencias de usuario y la mejora de la ciberseguridad se benefician enormemente de la IA en la nube. Sin embargo, este enfoque también presenta desafíos significativos. La seguridad y la privacidad de los datos son preocupaciones primordiales, ya que la transferencia y el almacenamiento de información sensible en la nube pueden ser vulnerables a ataques cibernéticos. Además, la latencia y la dependencia de una conexión a internet robusta pueden afectar el rendimiento de las aplicaciones en tiempo real. A pesar de estos retos, la IA en la nube sigue siendo una herramienta poderosa que permite a las

organizaciones innovar rápidamente y mantenerse competitivas en un mercado global en constante cambio.

IA en el Edge (en el borde)

La inteligencia artificial en el Edge, o Edge AI, está transformando la manera en que procesamos y analizamos datos al llevar la capacidad de cómputo directamente a los dispositivos y sensores en el borde de la red, en lugar de depender exclusivamente de centros de datos centralizados. Esta tecnología permite aplicaciones en tiempo real en una variedad de sectores, como la salud, donde los dispositivos médicos pueden monitorear y analizar signos vitales instantáneamente, o en la industria automotriz, donde los vehículos autónomos pueden tomar decisiones rápidas y seguras basadas en el análisis de datos locales. Sin embargo, la implementación de IA en el Edge presenta varios retos significativos. La limitación de recursos en términos de energía y capacidad de procesamiento en dispositivos Edge requiere el desarrollo de algoritmos más eficientes y modelos de IA optimizados. Además, garantizar la seguridad y privacidad de los datos procesados localmente es crucial, ya que estos dispositivos a menudo manejan información sensible. A pesar de estos desafíos, la Edge AI promete revolucionar la forma en que interactuamos con la tecnología, ofreciendo soluciones más rápidas, seguras y personalizadas.

La implementación de Edge AI ya se encuentra en varios dispositivos del día a día, desde Smartphones donde nos ayuda a incorporar IA en el modem 5G para potenciar la conexión, en funciones de cámara para mejorar la fotografía bajo condiciones menos favorables de luz o ruido, para permitir una experiencia excelente en la visualización de contenido. Otro ejemplo es; la implementación de Edge AI en los televisores inteligentes en lo que se refiere a control a través de gestos o voz, mejora en la calidad de imagen. Una de las implementaciones recientes es la incorporación de Edge AI en módems y enrutadores Wi-Fi 7, donde la IA ayuda a manejar el ancho de banda, asignar la potencia de la antena dependiendo de la distancia que se encuentra el dispositivo, como así también mantener una señal fluida mientras el usuario se encuentra en movimiento dentro de una casa u oficina.

IA Hibrida

La IA híbrida de nube y Edge representa una poderosa combinación que maximiza las capacidades de procesamiento y análisis de datos al distribuir las tareas entre la nube y los dispositivos en el borde. Esta arquitectura permite que las aplicaciones de IA aprovechen la potencia de la nube para tareas

intensivas en datos y aprendizaje profundo, mientras que los dispositivos Edge manejan el procesamiento en tiempo real y la toma de decisiones local. Por ejemplo, en el sector de la manufactura, los sensores Edge pueden monitorear equipos en tiempo real para detectar fallos inminentes, mientras que la nube analiza grandes volúmenes de datos históricos para optimizar el mantenimiento predictivo. En el ámbito de las ciudades inteligentes, las cámaras de vigilancia con IA en el Edge pueden identificar incidentes de seguridad instantáneamente, mientras que la nube gestiona el análisis a largo plazo de patrones de tráfico y comportamiento. Sin embargo, esta dualidad también presenta desafíos significativos. La sincronización y la coherencia de datos entre la nube y los dispositivos Edge pueden ser complicadas, y la latencia en la comunicación puede afectar el rendimiento en aplicaciones críticas. Además, la seguridad y la privacidad de los datos son preocupaciones clave, ya que los datos sensibles deben ser protegidos tanto en tránsito como en almacenamiento. A pesar de estos retos, la IA híbrida de nube y Edge ofrece un enfoque equilibrado y eficiente para desplegar soluciones de IA avanzadas en un mundo cada vez más conectado.

APLICACIONES ACTUALES DE LA IA

En el presente, las aplicaciones de la inteligencia artificial (IA) están transformando diversos sectores, mejorando la eficiencia, la precisión y la personalización de servicios y productos. En la salud, la IA se utiliza para diagnosticar enfermedades a través del análisis de imágenes médicas y datos de pacientes, permitiendo diagnósticos más rápidos y precisos. En el comercio electrónico, los algoritmos de recomendación impulsados por IA personalizan la experiencia de compra, sugiriendo productos basados en el comportamiento y las preferencias del usuario. En el sector financiero, la IA ayuda a detectar fraudes en tiempo real y a gestionar riesgos mediante el análisis de grandes volúmenes de datos transaccionales. Además, en la industria automotriz, los vehículos autónomos y los sistemas avanzados de asistencia al conductor (ADAS) utilizan IA para mejorar la seguridad y la eficiencia en la conducción. En el ámbito de la atención al cliente, los chatbots y asistentes virtuales, impulsados por IA, proporcionan soporte 24/7, resolviendo consultas y problemas de manera rápida y eficiente. Estas aplicaciones demuestran cómo la IA está integrándose en nuestra vida cotidiana, ofreciendo soluciones innovadoras y mejorando la calidad de los servicios en múltiples áreas.

II. EL PANORAMA DE LA IA EN DIFERENTES REGIONES DEL MUNDO

El panorama de la inteligencia artificial (IA) varía significativamente entre las diferentes regiones del mundo, influenciado por factores como la inversión en investigación, infraestructura tecnológica, políticas gubernamentales, acceso a datos, educación y el nivel de desarrollo económico. A continuación, ofrecemos un panorama general de cómo se está desarrollando la IA en distintas regiones del mundo:

I. AMÉRICA DEL NORTE (ESTADOS UNIDOS Y CANADÁ)

Estados Unidos es uno de los líderes globales en investigación y desarrollo de IA. El país alberga algunas de las principales empresas tecnológicas que impulsan la innovación en IA, como Google, Microsoft, Amazon, Apple, y muchas *startups*.

Además, universidades de renombre como el MIT, Stanford, y la Universidad de California en Berkeley son centros de investigación clave.

- Áreas de enfoque: IA aplicada a la automatización, inteligencia artificial generativa, aprendizaje automático, procesamiento de lenguaje natural, IA en la salud, robótica, vehículos autónomos e IA ética.
- Iniciativas gubernamentales: En 2021, Estados Unidos lanzó la "National AI Initiative Act", una estrategia para fomentar la investigación, el desarrollo y la adopción de IA en diversas industrias.
- Desafíos: El principal desafío es la regulación ética, la privacidad y la seguridad de los datos, así como la creciente competencia internacional en IA.

Canadá también juega un papel importante, con centros de investigación destacados en Toronto, Montreal y Vancouver. El gobierno canadiense ha impulsado políticas pro-IA y ha atraído inversiones para hacer del país un líder en IA ética y explicativa.

II. EUROPA

Europa se enfrenta a un panorama mixto en cuanto a la adopción de la IA, con diferentes niveles de desarrollo en sus países. Sin embargo, la Unión Europea (UE) ha liderado en cuanto a la regulación de la IA y la integración ética de la tecnología.

A. Áreas de enfoque: IA aplicada a la industria manufacturera, automotriz, servicios financieros, salud y agricultura. Europa también pone un fuerte énfasis en la IA explicativa y responsable.
B. Iniciativas gubernamentales: La UE tiene una estrategia de IA centrada en la ética y la sostenibilidad. En 2021, la Comisión Europea presentó un marco regulatorio para la IA que incluye leyes sobre la confianza, la transparencia y la protección de los derechos fundamentales.
C. Desafíos: A pesar de la inversión y la regulación, Europa enfrenta retos en términos de cohesión y falta de capital de riesgo en comparación con EE. UU. y China. Además, la falta de una infraestructura de datos común y el envejecimiento de la población pueden limitar el crecimiento en ciertos sectores.

Reino Unido es otro actor destacado, con grandes inversiones en IA en áreas como la salud, la automoción y la energía. Londres, Cambridge y Edimburgo son centros clave de innovación.

III. ASIA-PACÍFICO

Asia está viendo un crecimiento masivo en la adopción de IA, especialmente en China, pero también en Japón, Corea del Sur, Singapur e India.

China:
China ha superado a EE. UU. en términos de publicaciones científicas relacionadas con IA y tiene grandes ambiciones de convertirse en el líder mundial en IA para 2030. El gobierno chino ha hecho de la IA una prioridad nacional, con fuertes inversiones tanto en la investigación como en la adopción comercial de la tecnología.

- Áreas de enfoque: Reconocimiento facial, IA para vigilancia, vehículos autónomos, IA en la salud y fintech.
- Desafíos: Los problemas de privacidad, la censura y las preocupaciones sobre el control social mediante IA son temas controvertidos. Además, la guerra tecnológica con EE. UU. está impulsando la competencia en IA, lo que podría generar riesgos de desglobalización.

Japón:
Japón es un líder en robótica e inteligencia artificial aplicada a la manufactura y la automatización. Además, se ha enfocado en el uso de IA para ayudar a resolver problemas sociales, como el envejecimiento de la población.

- **Áreas de enfoque:** Robótica, IA en la automoción, IA en la atención médica y la vida diaria (como la robótica doméstica).
- **Desafíos:** El envejecimiento de la población japonesa y la escasez de mano de obra son impulsores clave de la adopción de IA, pero el país también se enfrenta a una falta de talento y recursos humanos para desarrollar la tecnología a gran escala.

Corea del Sur:
El gobierno surcoreano ha establecido planes de inversión en IA, destacando su enfoque en áreas como la fabricación inteligente y la IA aplicada al entretenimiento y los videojuegos.

- **Áreas de enfoque:** IA en la educación, la medicina, la cultura digital, el entretenimiento y la automoción.
- **Desafíos:** La necesidad de desarrollar una regulación ética de IA y la competencia con gigantes de IA como China y EE. UU.

India:
India tiene un panorama emergente de IA, centrado en su base de talento en ingeniería y en la adopción de soluciones tecnológicas en sectores como la salud, la agricultura y las finanzas.

- **Áreas de enfoque:** IA en la agricultura, IA en la educación, IA para la mejora de la calidad de vida en áreas rurales.
- **Desafíos:** Los desafíos económicos y sociales, como la desigualdad en el acceso a la tecnología y la falta de infraestructura digital en áreas rurales.

IV. AMÉRICA LATINA

En Latinoamérica, la adopción de IA está aún en una fase de crecimiento, pero se están viendo avances significativos, especialmente en países como Brasil, México, Argentina, Chile y Colombia.

- **Áreas de enfoque:** La IA se está utilizando principalmente en áreas como el sector bancario (fintech), la agricultura, la educación y la salud.
- **Iniciativas gubernamentales:** Varios países han comenzado a crear estrategias nacionales de IA. Brasil, por ejemplo, lanzó su estrategia nacional de IA en 2021, y Chile está trabajando en una serie de iniciativas para mejorar la adopción de IA en la economía digital

Desafíos:

- Infraestructura: Muchos países carecen de la infraestructura digital necesaria para una implementación masiva de IA.
- Brecha de talento: La falta de profesionales capacitados en IA sigue siendo un desafío, aunque el potencial de la región es grande, con un ecosistema de startups emergente y el uso de IA para resolver problemas sociales y económicos.
- Desigualdad digital: El acceso a la tecnología y a internet es desigual, lo que limita la expansión de soluciones basadas en IA en las zonas rurales y menos desarrolladas.

Propuesta de Agenda Digital para América Latina:

Con base en las tendencias actuales en materia de transformación digital, podríamos proponer los siguientes ejes de acción para la nueva etapa de la Agenda Digital para América Latina y el Caribe:

A. Fortalecer la Infraestructura Digital:
   - Expansión de la conectividad: Priorizar la conexión de zonas rurales y poblaciones de bajos ingresos.
   - Desarrollo de redes 5G: Impulsar la adopción de tecnologías de última generación para mejorar la velocidad y calidad de las conexiones.
   - Inversión en infraestructura de datos: Crear centros de datos regionales para almacenar y procesar grandes volúmenes de información.

B. Promover la Adopción de Tecnologías Digitales:
   - Capacitación digital: Implementar programas de formación para la población en general y para los trabajadores en particular.
   - Fomento de la innovación: Crear ecosistemas de innovación y emprendimiento digital.
   - Incentivos para la adopción de tecnologías: Ofrecer incentivos fiscales y financieros a las empresas que adopten soluciones digitales.

C. Desarrollar una Inteligencia Artificial responsable:
   - Establecer marcos regulatorios: Crear normativas claras y transparentes para el desarrollo y uso de la IA.
   - Invertir en investigación: Fomentar la investigación en IA y sus aplicaciones para el desarrollo.
   - Promover la ética en la IA: Asegurar que el desarrollo y uso de la IA se basen en principios éticos y respeten los derechos humanos.

D. Cerrar la Brecha Digital de Género:
   - Promover la participación de las mujeres: Fomentar la participación de las mujeres en el sector tecnológico.

- **Diseñar tecnologías inclusivas:** Asegurar que las tecnologías digitales sean diseñadas y desarrolladas de manera inclusiva.

E. Cooperación Regional y Gobernanza Digital:

- **Fortalecer los mecanismos de cooperación:** Reforzar los mecanismos de coordinación y colaboración entre los países de la región.
- **Desarrollar capacidades institucionales:** Fortalecer las capacidades de los gobiernos para diseñar e implementar políticas públicas en materia digital.
- **Promover la gobernanza digital:** Establecer principios comunes de gobernanza digital a nivel regional.

F. Seguridad Cibernética:

- **Invertir en ciberseguridad:** Fortalecer las capacidades de los países en materia de ciberseguridad.
- **Cooperación internacional:** Establecer mecanismos de cooperación internacional para combatir las amenazas cibernéticas.

V. ÁFRICA

El panorama de la IA en África es emergente, y aunque la región enfrenta desafíos significativos en términos de infraestructura y capital humano, hay áreas en las que se está haciendo un avance notable.

Áreas de enfoque:

- **Agricultura**: La IA se usa para mejorar la producción agrícola mediante predicción del clima y soluciones para la gestión de cultivos.
- **Salud**: El uso de la IA para diagnosticar enfermedades y mejorar los sistemas de atención médica es clave.
- **Fintech**: En algunos países como Kenia y Nigeria, las aplicaciones de IA en el sector financiero están ayudando a mejorar el acceso a servicios bancarios.

Desafíos:

- **Falta de infraestructura**: La infraestructura de Internet y de datos es limitada en muchas partes de África.
- **Brecha de habilidades**: La falta de educación y formación especializada en IA es un reto importante para el crecimiento de la tecnología en el continente.

El panorama global de la IA está muy influenciado por los recursos económicos, las políticas gubernamentales, la infraestructura tecnológica, la educación y la disponibilidad de talento en cada región. Mientras que países como Estados Unidos, China y algunos países de Europa y Asia están a la vanguardia de la investigación y la aplicación de IA, otras regiones como África y América Latina están en una fase más temprana de adopción, enfrentando desafíos específicos en términos de infraestructura, capital humano y acceso a la tecnología. Sin embargo, todas las regiones están mostrando un creciente interés en la inteligencia artificial, reconociendo su potencial para transformar la economía global y abordar problemas sociales y ambientales.

Principales Actores y Ecosistema de IA

El ecosistema de Inteligencia Artificial (IA) ha evolucionado de manera espectacular en los últimos años, y actualmente está siendo transformado por una convergencia de actores globales, proveedores de servicios en la nube, empresas de redes neuronales, y fabricantes de hardware especializado, como los chips para IA. Todos ellos están impulsando la capacidad de ejecutar IA en el borde, es decir, cerca de la fuente de los datos, en lugar de depender únicamente de la nube. Esta tendencia busca mejorar la velocidad, reducir la latencia, y preservar la seguridad de la información de los individuos.

Proveedores de Servicios en la Nube

Los gigantes tecnológicos en la nube ofrecen plataformas robustas para IA, proporcionando las infraestructuras necesarias para el procesamiento, almacenamiento y análisis de grandes volúmenes de datos. Estos proveedores ofrecen soluciones tanto para el procesamiento en la nube como para el despliegue de IA en el borde.

- Amazon Web Services (AWS): AWS es un actor clave en la provisión de servicios para IA, ofreciendo herramientas como Amazon SageMaker, que permite a los desarrolladores crear, entrenar e implementar modelos de IA. AWS también está liderando el camino con su infraestructura de borde, ofreciendo servicios como AWS IoT Greengrass, que permite ejecutar IA en dispositivos en el borde sin necesidad de enviar todos los datos a la nube.
- Google Cloud: Google es un líder en inteligencia artificial, con su plataforma Google Cloud AI que abarca desde el entrenamiento de modelos hasta la implementación en producción. Google también ofrece soluciones para IA en el borde, como Google Edge TPU y TensorFlow Lite, que permiten ejecutar modelos de IA directamente en dispositivos locales.

- Microsoft Azure: Microsoft Azure se ha posicionado como un proveedor esencial en el ecosistema de IA, con herramientas como Azure Machine Learning y su capacidad para implementar IA en el borde a través de Azure IoT Edge, que permite ejecutar modelos de IA directamente en dispositivos conectados, manteniendo la eficiencia y la seguridad.

Redes Neuronales Más Grandes

Las redes neuronales están en el centro de la revolución de la IA, con modelos cada vez más grandes y sofisticados. Algunas de las redes neuronales más avanzadas están diseñadas para realizar tareas como el procesamiento del lenguaje natural (PLN), la visión por computadora, la clasificación de datos y mucho más.

- OpenAI: El desarrollo de modelos como GPT (Generative Pre-trained Transformer) ha impulsado la IA generativa, una de las áreas de mayor crecimiento en IA. OpenAI se ha enfocado en la creación de modelos de lenguaje que son capaces de realizar tareas complejas con un alto nivel de comprensión.
- DeepMind (Alphabet): DeepMind, una subsidiaria de Google es conocida por desarrollar redes neuronales avanzadas que han logrado hitos como la creación de AlphaGo, un sistema que derrotó a campeones humanos en el juego de Go. DeepMind continúa siendo pionera en el desarrollo de IA generalista.
- Meta (Facebook): Meta ha invertido fuertemente en el desarrollo de redes neuronales para el procesamiento de texto, imágenes y video. Su modelo
- LLaMA (Large Language Model Meta AI) es uno de los más avanzados en la actualidad, con un enfoque en la creación de modelos de código abierto.

Fabricantes de Chipsets (Ejemplo: MediaTek)

Los fabricantes de chips desempeñan un papel fundamental al proporcionar el hardware necesario para ejecutar modelos de IA de manera eficiente, especialmente en dispositivos de borde. Estos chips están diseñados para manejar la inferencia de modelos de IA con baja latencia, sin necesidad de depender de la nube.

- MediaTek: MediaTek ha sido un jugador importante en el mercado de chips para dispositivos móviles y de borde. Sus chipsets, como la serie Dimensity, están diseñados para ejecutar tareas de IA de forma eficiente en dispositivos móviles, IoT, y otros dispositivos conectados. La serie Dimensity incorpora unidades de procesamiento de inteligencia artificial (NPU) que permiten la

ejecución de modelos de IA en el dispositivo, lo que reduce la necesidad de enviar datos a la nube y mejora la privacidad y seguridad.

- NVIDIA: NVIDIA es conocida por su liderazgo en el desarrollo de procesadores gráficos (GPU) especializados en el entrenamiento de IA, pero también ha avanzado en el hardware para la inferencia en el borde con productos como Jetson y Tegra, que permiten ejecutar modelos de IA en dispositivos locales con alta eficiencia energética.

La IA al Borde: Seguridad y Privacidad

La tendencia de ejecutar IA en el borde no solo mejora la eficiencia y reduce la latencia, sino que también tiene un impacto positivo en la seguridad y privacidad de los datos. Algunos de los beneficios más importantes de esta tendencia incluyen:

- Minimización de la transferencia de datos: Al procesar los datos localmente en el dispositivo o en la red de borde, se reduce la necesidad de enviar información sensible a la nube, lo que minimiza los riesgos de exposición a posibles violaciones de seguridad.
- Control de datos por parte del usuario: Los usuarios pueden mantener el control sobre sus datos, lo que es especialmente importante en un contexto donde la privacidad de la información personal es un tema cada vez más relevante.
- Desempeño local con privacidad reforzada: Al realizar la inferencia de IA de manera local en el dispositivo, se asegura que los datos no salgan del dispositivo, lo que proporciona una capa adicional de seguridad al evitar la transmisión de datos personales a servidores externos.

El ecosistema de IA está experimentando una transformación significativa, y los principales actores globales están desempeñando un papel clave en habilitar la IA al borde. Empresas como AWS, Google, Microsoft, y MediaTek están impulsando el desarrollo de soluciones que permiten ejecutar IA de manera más eficiente, rápida y segura, preservando al mismo tiempo la privacidad de los datos. En este contexto, la IA al borde representa no solo una mejora en el rendimiento, sino también un avance crucial en la protección de la información personal de los individuos.

Sectores clave y aplicaciones de la IA

La Inteligencia Artificial (IA) está transformando diversos sectores mediante el desarrollo de soluciones avanzadas que mejoran la eficiencia, la toma de

decisiones y la automatización de procesos. A continuación, se describen algunos de los **sectores clave** donde la IA está teniendo un impacto significativo, junto con sus principales **aplicaciones**:

I. Salud

La IA está revolucionando la atención médica en diversas áreas, como el diagnóstico, la personalización del tratamiento y la investigación.

- **Diagnóstico asistido**: Herramientas de IA como los sistemas de reconocimiento de imágenes médicas (radiografías, resonancias magnéticas, etc.) ayudan a detectar enfermedades como el cáncer, enfermedades cardiovasculares, y problemas neurológicos con gran precisión.
- **Medicina personalizada**: Algoritmos de IA pueden analizar datos genéticos y de comportamiento para diseñar tratamientos más específicos y efectivos.
- **Robots quirúrgicos**: La cirugía asistida por robots, que se guía por IA, permite procedimientos más precisos y menos invasivos.
- **Asistentes virtuales de salud**: Aplicaciones como chatbots de salud ofrecen asesoramiento médico preliminar o seguimiento de pacientes a través de plataformas digitales.

II. Automotriz

La IA está desempeñando un papel crucial en la mejora de la seguridad, la eficiencia y la experiencia de conducción.

- **Vehículos autónomos**: Los autos sin conductor utilizan IA para procesar información en tiempo real y tomar decisiones sobre navegación, seguridad y control del vehículo.
- **Sistemas de asistencia al conductor (ADAS)**: Asistentes como el frenado automático, el mantenimiento de carril, y la detección de obstáculos dependen de la IA para mejorar la seguridad vial.
- **Optimización de la fabricación**: La IA se usa en la producción de vehículos para automatizar procesos de ensamblaje, realizar mantenimiento predictivo y optimizar la cadena de suministro.

III. Finanzas

La IA está cambiando la manera en que las instituciones financieras gestionan el riesgo, los servicios al cliente y la toma de decisiones estratégicas.

- **Análisis predictivo:** La IA puede predecir tendencias del mercado, el comportamiento de los consumidores y otros factores económicos, lo que ayuda a las instituciones a tomar decisiones informadas.

- Detección de fraudes: Los algoritmos de IA monitorean transacciones en tiempo real para identificar patrones sospechosos y prevenir fraudes.
- Asistentes virtuales y robo-advisors: Los chatbots financieros y los asesores automáticos utilizan IA para ofrecer recomendaciones personalizadas en inversiones, seguros y otros productos financieros.
- Trading algorítmico: Los algoritmos de IA pueden ejecutar operaciones bursátiles de manera más rápida y precisa que los humanos, basándose en grandes volúmenes de datos.

IV. Comercio electrónico y marketing

La IA está optimizando la experiencia del cliente, la personalización y las estrategias de ventas en línea.

- Recomendaciones personalizadas: Plataformas como Amazon o Netflix utilizan IA para recomendar productos o contenidos según el comportamiento y las preferencias de los usuarios.
- Chatbots de atención al cliente: Los chatbots impulsados por IA pueden manejar consultas de clientes, resolver problemas y proporcionar soporte 24/7.
- Análisis de datos y predicción de tendencias: Herramientas de IA analizan grandes volúmenes de datos para entender las preferencias de los consumidores y prever comportamientos futuros, lo que permite una segmentación de mercado más precisa.
- Automatización del marketing: Los algoritmos de IA permiten crear campañas publicitarias personalizadas basadas en el comportamiento de los usuarios, optimizando así el ROI.

V. Industria y manufactura

La IA está mejorando la eficiencia, la calidad y la seguridad en los procesos industriales.

- Mantenimiento predictivo: Sensores inteligentes y algoritmos de IA analizan el estado de las máquinas y equipos para predecir fallos antes de que ocurran, evitando tiempos de inactividad costosos.
- Optimización de la cadena de suministro: La IA mejora la logística y la gestión de inventarios mediante el análisis de datos en tiempo real, lo que reduce los costos y mejora la eficiencia operativa.
- Robots industriales: La automatización de tareas repetitivas mediante robots autónomos o colaborativos (cobots) mejora la precisión y reduce la necesidad de intervención humana.

VI. Educación

La IA está transformando la forma en que enseñamos y aprendemos, brindando herramientas más personalizadas y eficientes.

- Sistemas de tutoría inteligente: Plataformas de aprendizaje adaptativo utilizan IA para personalizar el contenido y el ritmo de los cursos según las necesidades de cada estudiante.
- Evaluación automatizada: Los algoritmos de IA pueden calificar exámenes y tareas, brindando retroalimentación rápida y precisa.
- Análisis del rendimiento: Los sistemas de IA pueden identificar patrones en el rendimiento de los estudiantes, ayudando a los educadores a intervenir de manera más efectiva.

VII. Energía y sostenibilidad

La IA está ayudando a mejorar la eficiencia energética y a optimizar la distribución de recursos naturales.

- Optimización de redes eléctricas: Los algoritmos de IA gestionan las redes de distribución de energía para garantizar un suministro eficiente y reducir pérdidas.
- Energías renovables: La IA ayuda en la predicción y optimización de la generación de energía a partir de fuentes renovables, como solar y eólica, ajustando la distribución de energía de acuerdo con las condiciones climáticas.
- Gestión del consumo energético: Las casas y edificios inteligentes utilizan IA para controlar el consumo energético, mejorando la eficiencia y reduciendo costos.

VIII. Agricultura

La IA está mejorando la producción agrícola mediante el análisis de datos y la automatización de procesos.

- Agricultura de precisión: Sensores y drones equipados con IA recogen datos sobre el estado del suelo y los cultivos, permitiendo un uso más eficiente de recursos como agua y fertilizantes.
- Robots agrícolas: La automatización de tareas como la cosecha, la siembra y el monitoreo de cultivos reduce la necesidad de trabajo manual y mejora la eficiencia.
- Predicción de rendimientos: Los modelos predictivos basados en IA pueden prever el rendimiento de los cultivos, lo que ayuda a los agricultores a planificar mejor la cosecha y la distribución.

IX. Seguridad y vigilancia

La IA también juega un papel importante en la mejora de la seguridad en entornos físicos y digitales.

- Reconocimiento facial y de objetos: Las cámaras de vigilancia equipadas con IA pueden identificar personas y objetos en tiempo real, mejorando la seguridad pública y privada.
- Análisis forense: La IA facilita la recopilación y análisis de grandes volúmenes de datos para resolver crímenes y mejorar la investigación.
- Ciberseguridad: Algoritmos de IA detectan y responden a amenazas cibernéticas de manera más rápida y precisa, protegiendo redes, sistemas y datos sensibles.

NIVELES DE ADOPCIÓN DE LA INTELIGENCIA ARTIFICIAL EN AMÉRICA LATINA

La brecha en gasto en IA en América Latina es un desafío importante, pero también una oportunidad para impulsar el desarrollo de la región. Es fundamental que los gobiernos, las empresas y la academia trabajen en conjunto para fomentar la inversión en esta tecnología y aprovechar todo su potencial.

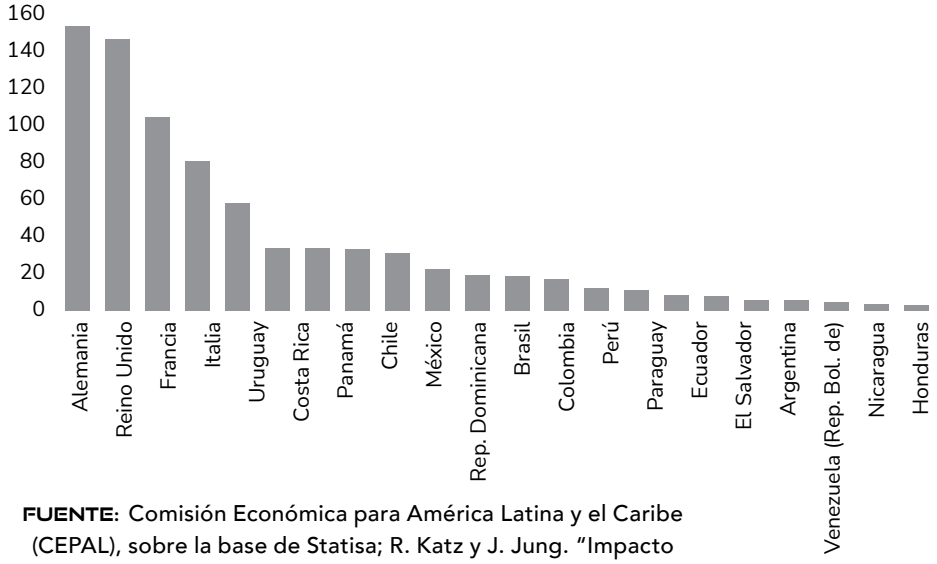
Existe una clara disparidad en la adopción de herramientas de inteligencia artificial entre Europa y América Latina. Los países nórdicos, como Dinamarca y Suecia, lideran la adopción a nivel mundial, mientras que los países latinoamericanos se encuentran notablemente rezagados.

FACTORES CLAVE QUE EXPLICAN LA BRECHA:

I. Infraestructura Tecnológica:

- Conectividad a Internet: Una infraestructura de internet robusta y de alta velocidad es esencial para el desarrollo y uso de la IA. Los países europeos, en general, cuentan con una mayor penetración de internet y velocidades de conexión más altas.
- Acceso a Computación en la Nube: La nube proporciona los recursos computacionales necesarios para entrenar y ejecutar modelos de IA. La disponibilidad y asequibilidad de servicios en la nube pueden variar significativamente entre regiones.

GRÁFICA 1 Gasto en Inteligencia artificial per cápita, 2023 (en dólares). América Latina (18 países) y Europa (4 países)



FUENTE: Comisión Económica para América Latina y el Caribe (CEPAL), sobre la base de Statisa; R. Katz y J. Jung. "Impacto económico de la inteligencia artificial en América Latina", Santiago, Telecom Advisory Services/Comisión Económica para América Latina y el Caribe (CEPAL), 2024, inédito.

NOTA El gasto se midió sobre la base del ingreso de los proveedores de inteligencia artificial y las estimaciones se basaron en el ingreso promedio por usuario.

II. Capital Humano:

- Educación y Capacitación: La formación en áreas como ciencia de datos, ingeniería de software y matemáticas es fundamental para el desarrollo y la implementación de soluciones de IA. Los países europeos suelen invertir más en educación superior y en programas de capacitación en tecnologías emergentes.
- Talento Especializado: La disponibilidad de profesionales con las habilidades necesarias para trabajar con IA es otro factor crítico. La fuga de cerebros hacia países desarrollados puede limitar el talento disponible en algunos países latinoamericanos.

III. Marco Regulatorio:

- Políticas Públicas: Las políticas gubernamentales pueden fomentar o restringir la adopción de la IA. Los marcos regulatorios claros y favorables pueden incentivar la inversión y la innovación en este campo.

- Protección de Datos: Las regulaciones sobre protección de datos, como el RGPD en Europa, pueden influir en la manera en que las empresas recolectan, almacenan y utilizan datos para desarrollar modelos de IA.

IV. Madurez del Mercado:

- Demanda de Soluciones de IA: La demanda de soluciones de IA por parte de las empresas y los consumidores es un motor clave para la adopción. Los mercados más maduros, como los europeos, suelen tener una mayor demanda de productos y servicios basados en IA.
- Ecosistema de Innovación: La existencia de un ecosistema de innovación vibrante, con startups, aceleradoras y fondos de inversión, puede acelerar la adopción de nuevas tecnologías.

A MANERA DE RESUMEN

La IA está evolucionando rápidamente y tiene aplicaciones en casi todos los sectores imaginables. A medida que la tecnología continúa avanzando, se espera que su impacto sea aún más profundo, ofreciendo soluciones innovadoras que transformen la forma en que interactuamos con el mundo. Sin embargo, también es crucial abordar los desafíos éticos, legales y sociales asociados con la adopción de la IA, para garantizar que sus beneficios sean equitativos y sostenibles.

III. DESAFÍOS TECNOLÓGICOS.

La inteligencia artificial (IA) está revolucionando múltiples sectores, desde la salud y la educación hasta las finanzas y la agricultura. Sin embargo, la implementación efectiva de IA enfrenta una serie de desafíos tecnológicos de infraestructura que deben ser abordados para aprovechar plenamente su potencial. Estos desafíos incluyen la necesidad de una infraestructura de datos robusta, la capacidad de procesamiento, la conectividad y la seguridad.

INFRAESTRUCTURA DE DATOS

Los sistemas de IA dependen en gran medida de grandes volúmenes de datos de alta calidad para su entrenamiento y operación. La gestión de estos datos implica varios desafíos:

- **Recolección e Integración de Datos:** Agregar datos de diversas fuentes y asegurarse de que estén en un formato utilizable puede ser complejo y llevar mucho tiempo.
- **Limpieza y Preprocesamiento de Datos:** Los datos en bruto a menudo contienen ruido, valores faltantes e inconsistencias que deben abordarse antes de que puedan usarse de manera efectiva.
- **Almacenamiento de Datos:** Almacenar grandes cantidades de datos requiere soluciones de almacenamiento escalables y eficientes que puedan manejar tanto datos estructurados como no estructurados.

Todo esto presentan los desafíos en la infraestructura de datos. Sin embargo, muchas organizaciones en el mundo y América Latina carecen de sistemas adecuados para recopilar, almacenar y gestionar estos datos. La fragmentación de datos, la falta de estandarización y la calidad inconsistente de los datos son problemas comunes que pueden afectar negativamente el rendimiento de los sistemas de IA. Además, la integración de datos de múltiples fuentes y formatos requiere soluciones avanzadas de gestión de datos y plataformas de interoperabilidad que aún se encuentran en desarrollo.

CAPACIDAD DE PROCESAMIENTO

La IA requiere una capacidad de procesamiento significativa para manejar algoritmos complejos y grandes conjuntos de datos. Esto implica la necesidad de hardware especializado potente, como unidades de procesamiento gráfico (GPU) y unidades de procesamiento tensorial (TPU) y Unidades de Procesamiento Neuronales (NPU), que son capaces de realizar cálculos paralelos a alta velocidad. Sin embargo, la adquisición y mantenimiento de este hardware puede ser costoso, especialmente para pequeñas y medianas empresas. Además, la escalabilidad de la infraestructura de procesamiento es crucial para adaptarse a las crecientes demandas de las aplicaciones de IA.

Requisitos de Memoria: Los modelos y conjuntos de datos grandes requieren una capacidad de memoria significativa para almacenar y procesar datos de manera eficiente.

COMPLEJIDAD DE LOS ALGORITMOS

La complejidad de los algoritmos de IA presenta varios desafíos computacionales:

- **Tiempo de Entrenamiento:** Entrenar modelos sofisticados puede llevar días o incluso semanas, dependiendo del tamaño del conjunto de datos y la complejidad del modelo.
- **Ajuste de Hiperparámetros:** Encontrar los hiperparámetros óptimos para un modelo a menudo implica una extensa experimentación y esfuerzo computacional.
- **Optimización del Modelo:** Asegurar que los modelos estén optimizados para el rendimiento y la precisión requiere técnicas avanzadas y recursos computacionales

CONECTIVIDAD

La conectividad es otro desafío crítico. La implementación de IA a menudo requiere una conectividad de red rápida y confiable para transferir grandes volúmenes de datos entre dispositivos y sistemas. En muchas regiones, especialmente en áreas rurales y en desarrollo, la infraestructura de telecomunicaciones aun es insuficiente para soportar estas demandas. La falta de conectividad adecuada puede limitar la capacidad de las organizaciones para implementar soluciones de IA en tiempo real y aprovechar tecnologías emergentes como el Internet de las Cosas (IoT).

Muchas aplicaciones de IA, como los vehículos autónomos y el análisis en tiempo real, requieren procesamiento y toma de decisiones inmediatos:

- **Latencia:** Minimizar la latencia es crucial para las aplicaciones en tiempo real para asegurar respuestas oportunas y precisas.
- **Computación en el Borde (Edge Computing):** Desplegar modelos de IA en dispositivos edge, como sensores y dispositivos móviles, requiere algoritmos eficientes que puedan operar con recursos computacionales y energía limitados.

SEGURIDAD Y PRIVACIDAD

La seguridad y la privacidad de los datos son preocupaciones fundamentales en la implementación de IA. Los sistemas de IA a menudo manejan información sensible y personal, lo que los convierte en objetivos atractivos para ciberataques. La protección de estos datos requiere medidas de seguridad robustas, como el cifrado, la autenticación multifactor y la detección de intrusiones. Además,

las organizaciones deben cumplir con regulaciones de privacidad de datos, como el Reglamento General de Protección de Datos (GDPR) en Europa y leyes similares en otras regiones. La implementación de estas medidas puede ser compleja y costosa, pero es esencial para mantener la confianza de los usuarios y proteger la integridad de los sistemas de IA.

ENERGÍA Y SOSTENIBILIDAD

La infraestructura de IA también plantea desafíos en términos de consumo de energía y sostenibilidad. Los centros de datos y el hardware especializado para IA consumen grandes cantidades de energía, lo que puede tener un impacto ambiental significativo. Las organizaciones deben buscar soluciones sostenibles, como el uso de energías renovables y la optimización de la eficiencia energética, para mitigar este impacto.

Además, la gestión eficiente de los recursos y la reducción de la huella de carbono son consideraciones importantes para el desarrollo sostenible de la infraestructura de IA.

A MANERA DE RESUMEN

La implementación de inteligencia artificial enfrenta una serie de desafíos tecnológicos de infraestructura que deben ser abordados para aprovechar plenamente su potencial. La infraestructura de datos, la capacidad de procesamiento, la conectividad, la seguridad y la sostenibilidad son áreas clave que requieren atención e inversión. Superar estos desafíos no solo permitirá a las organizaciones implementar soluciones de IA de manera efectiva, sino que también impulsará la innovación y el crecimiento en múltiples sectores. A medida que la tecnología avanza, es crucial que las organizaciones y los gobiernos trabajen juntos para desarrollar una infraestructura robusta y sostenible que soporte el futuro de la inteligencia artificial.

IV. DESAFÍOS SOCIALES Y CULTURALES

ACEPTACIÓN Y PERCEPCIÓN PÚBLICA DE LA IA

La IA ha tenido una aceptación y adopción masiva en la última década, con el advenimiento de los asistentes activados por voz, teléfonos inteligentes (Siri y asistente de Google) y tabletas para acceder a internet. Por lo que podemos

inferir que la IA ha sido adoptada mayormente por quienes hoy poseen un smartphone por ser el dispositivo más popular. En América Latina siendo es el dispositivo más usado para acceder a Internet, así como de los primeros aplicativos y funciones masivas de IA. De acuerdo con el reporte de la GSMA sobre la economía móvil en América Latina 2024, los usuarios de internet móvil llegan a 418 millones en con una tasa de penetración del 65%, así como estima que se alcancen 485 millones en 2030 con una tasa de penetración del 72%. En México, 97.2 millones de personas usan el teléfono celular, lo que representa 81.4% de la población según el reporte del 2023, publicado el 13 de junio de 2024 sobre la encuesta nacional sobre disponibilidad y uso de tecnología de la información los hogares (ENDUTIH) del INEGI.

El reconocimiento de objetos y reconocimiento facial han sido de las funciones pioneras de IA y que cada vez más reemplaza el uso de las claves para acceder y desbloquear los dispositivos, aplicativos de uso privado como son los servicios bancarios, detección de personas y objetos para cámaras de videovigilancia, etc. En aplicaciones multimedia como en fotografía o video, la IA ayuda a mejorar la captura y despliegue de información capturada a través de la lente, y la IA generativa para hacer edición de contenido. Así como muchos otros dispositivos domésticos que realizan sus tareas como robots para limpieza, bocinas inteligentes y demás dispositivos IoT inteligentes usados en forma cotidiana que inadvertidamente la IA se encuentran embebida. Todos estos dispositivos electrónicos son una realidad gracias a la evolución tecnológica y al desarrollo de semiconductores que mediante la integración en sistemas en chip (chipset) más robustos, que incorporan potentes procesadores de cómputo (CPU), procesadores gráficos rápidos (GPU) y de gran capacidad, añadiendo procesadores de IA (NPU) con mayor desempeño, para acelerar el procesamiento de los complejos algoritmos de IA. El resultado del trabajo en conjunto de estos tres procesadores hace posible que los dispositivos sus funciones con IA las realicen en tiempo real. El avance tecnológico de los últimos años ha hecho posible disminuir progresivamente el tamaño de los nodos en semiconductores y que actualmente hayan alcanzado hasta 3 nanómetros. Lo que hace una realidad poder diseñar complejos sistemas en chip en un encapsulado de aproximadamente 1 cm2 con billones de transistores y con mayor eficiencia de consumo de energía para fabricar dispositivos con IA.

El debut de Inteligencia Artificial Generativa (IAG) en el mercado occidental ocurrió con la llegada Chat GPT-3, Meta AI y Gemini, cuyo procesamiento se realiza en la nube, es decir que el contenido que genera para responder a la solicitud de usuario es procesado en servidores con gran capacidad tanto

cómputo, como de amplio manejo de memoria. En Asia se han desarrollado modelos similares como 0.1AI Yi, Tongyi, Baichuan AI, Hunyuan.

Podemos decir que actualmente nos encontramos en el proceso de adopción de la IAG y aprendiendo lo que es posible hacer. Es decir, sobre el tipo de contenido que podemos generar a partir de una solicitud de información, ya sea a través de texto, sonido, imagen o código; así como el tipo de formato en que deseamos el resultado. Por ejemplo, solicitar mediante texto que nos genere una imagen (Stable Diffussion), traducir un texto o mensaje de voz a otro idioma y el resultado en voz o texto o viceversa.

El procesamiento de la inteligencia artificial generativa demanda de mayores recursos que la IA analítica o predictiva. Para brindar una ligera idea, se requiere de más de diez veces la capacidad de cómputo en TOPs, y más de cien veces mayor capacidad de memoria y en general podemos decir que para lograr un resultado con mayor precisión se requiere de mayores recursos.

Los resultados de la IAG pueden ser no exactos y generar resultados imprecisos. La precisión depende de la cantidad de parámetros del modelo. Siendo de cientos de billones de parámetros los más exactos. Así como también si el origen de los datos de entrenamiento fue tomado de la internet, cabe la posibilidad de que el resultado contenga información que posea propiedad intelectual. Por lo que se debe tener cuidado al usar modelos en la nube con información que sea sensible de confidencialidad y/o propiedad intelectual, ya que ésta puede ser usada para responder a otro usuario que requiera de información similar.

Una de las aplicaciones más usadas en la industria de servicio son los conocidos chats bots, ya sean a través de texto o por voz, que en etapa temprana de desarrollo son enfadosos para los usuarios por su imperfección y falta de aprendizaje para brindar respuestas más asertivas. Sin embargo, su perfeccionamiento nos hará asumir que estamos conversando con una persona.

Existen diversos usos de la IAG en la industria que ayudan a mejorar la calidad o para disminuir el tiempo de desarrollo de nuevos productos. Sin embargo, es necesario probar los resultados generados por la IAG para conocer su efectividad antes de ponerlo en producción y con ello garantizar el éxito. De manera interesante, los modelos de IAG se perfeccionan, cuando retroalimentamos al modelo con los resultados reales que obtenemos al realizar la prueba, ya que el modelo aprende y se vuelve más inteligente.

Actualmente, existen dispositivos electrónicos con IAG "Edge" (en el borde, tratado en el capítulo 1) como son smartphones, computadoras, tabletas y otros dispositivos inteligentes IoT con aplicativos o funciones IAG, es decir que se

procesan y generan el resultado en el mismo dispositivo sin hacer uso de servicios en la nube, y consecuentemente lograr la preservación de información sensible de datos personales y/o confidenciales, tales como edición de fotografías y videos, resumen de textos o conversaciones, generación de imágenes y video, traducción simultánea de textos y conversaciones, por mencionar algunas de ellas. Esto es posible gracias a la innovación tecnológica en semiconductores y al desarrollo de sistemas en chip más complejos como se mencionó anteriormente, adicionando mayor capacidad de procesamiento e integrando múltiples modelos de IAG optimizados, así como de modelos de adaptación para optimizar los grandes modelos de IAG y lograr resultados con razonable precisión.

IMPACTO EN EL EMPLEO Y EL MERCADO LABORAL

Conceptualmente, el advenimiento de la Inteligencia Artificial puede considerarse como la tercera revolución industrial, ya que su adopción en la industria tendrá un impacto en la vida laboral y quizá en mayor medida con la IAG.

El uso de la IA ofrece ventajas con impacto positivo en el producto, al empresario, al trabajador y al consumidor. El uso de la IA en la industria ayuda a mejorar la calidad de los productos, ya que es posible detectar el origen de defectos y mediante la IAG aprender como corregirlos, mejorando los procesos de manufactura. Encontramos casos de uso en la cadena de distribución y logística, para mejorar el manejo de inventario, optimizar la compra de materia prima programada de acuerdo con las órdenes de clientes y pronósticos de ventas, y ligada a la administración, ayuda a mejorar la negociación de precios y programación de pagos, cobranza, e inversión de ingresos. Las mejoras de calidad y la optimización de procesos ayudan en la industria a ser más competitivos, ofreciendo mejores productos y optimizar el costo e incrementar la rentabilidad para los accionistas y generar espacio para ofrecer mejores condiciones laborales a sus trabajadores. El cliente se beneficia al adquirir productos de calidad que satisfagan sus necesidades a precio competitivo.

En la vida laboral, los trabajadores se benefician por la mitigación de riesgos y accidentes, ya que las máquinas o robots con IA realizan los trabajos en las áreas cuyas condiciones extremas pueden afectar la salud como son temperatura y ruido excesivos, inhalación y exposición a sustancias químicas (líquidos, gases y polvo) nocivos, así evitar laborar en sitios de alto peligro. El uso de la IA en general disminuye la actividad física y hasta la fatiga mental del trabajador, ya que reduce el trabajo repetitivo que tras horas de trabajo el cansancio genera errores.

El consumidor al comprar mejores productos libres de defectos los podrá disfrutar por más tiempo y reducir gastos de reparación y mantenimiento, hasta evitar accidentes como puede ser en la industria automotriz.

La adopción de IA en la industria no es un camino sencillo, ya que requiere de una buena planeación, hacer reingeniería de procesos, inversión de nueva tecnología, creación de nuevas posiciones de trabajo capacitación continua, cambio de hábitos y cultura de trabajo que pueden llevar a la resistencia de algunos colaboradores.

El uso de la IA en la industria trae consigo grandes tanto para la empresa como en las tareas del trabajador, implica educación en nuevas funciones y desarrollar nuevas habilidades, aprovechar la experiencia para generar información útil para entrenar los modelos de IA y etiquetarlos correctamente para obtener mejores resultados.

EDUCACIÓN Y CAPACITACIÓN EN IA

El campo de la Inteligencia Artificial es muy extenso y multidisciplinario. Por lo que solo enunciaremos algunos temas y recomendaciones para dominar su aprendizaje:

- Comprender los términos y conceptos básicos, conocer de programación y aprender algún lenguaje como Phyton usado para IA, probabilidad y estadística, cálculo y algebra lineal, tener los fundamentos de aprendizaje automático supervisado/no supervisado y sus diferencias y visión de algoritmos más comunes,
- Manejar proyectos sencillos de aprendizaje automático, básicos de aprendizaje profundo y redes neuronales, fundamentos de arquitectura de redes neuronales, marcos de trabajo de aprendizaje profundo como TensorFlow y Pytorch, construir y entrenar una red neuronal, conceptos de sobreajuste/bajo ajuste, preprocesamiento de datos y datos de ingeniería, Evaluar modelos de aprendizaje automático efectivos, arquitecturas de redes neuronales convolutivas y recurrentes, trabajar un proyecto de visión computarizada.
- Básicos de procesamiento de lenguaje natural, construir un proceso de lenguaje natural, revisión de los algoritmos y el refuerzo de aprendizaje, elaborar un proyecto de aprendizaje por refuerzo básico, conceptos de modelos generativos, construir un modelo generativo, estudiar consideraciones éticas de IA, desarrollar aplicaciones de IA reales, conocer sobre servicios de IA basados en la nube, desarrollar un modelo de IA en la nube.

- Conocer la importancia de la IA en diferentes industrias, estudiar algoritmos de IA par tareas, tecnologías de grandes datos e IA, análisis de series de tiempo con IA, técnicas de optimización de modelos de IA, modelos de transferencia de enseñanza y preentrenamiento, investigación de trabajos de IA, contribuir en proyectos de IA de código abierto, construir un portafolio de IA.
- Explorar tópicos avanzados de meta aprendizaje, las últimas tendencias y avances de IA, relacionarse con comunidades de IA, certificarse en AI avanzada, estudiar libros y artículos sobre IA, actualizarse en lo blogs de noticias de IA, especialización de subcampos de IA, trabajar en proyectos de IA interdisciplinarios.

Actualmente existen un gran número de empresas que han desarrollado gran experiencia en IA, así como también diversos modelos entrenados de lenguaje amplio LLM, para ser adaptados a diferentes proyectos. Por lo que no es necesario ser un experto para implementar IA, sino comprender los conceptos básicos, y en algunos casos con ayuda de expertos aprender a generar información valiosa y hacer el etiquetado

DIVERSIDAD E INCLUSIÓN EN EL DESARROLLO DE IA

La IAG ayuda a generar material didáctico más rico en contenido multimedia voz, imágenes y video 3D, AR/VR; así como desarrollar mejor contenido, así como también facilita el diagnóstico en el aprendizaje.

Hay un futuro muy prometedor en la educación mediante el uso de la IA como un recurso para la enseñanza y comunicación adaptiva no solo para diferentes idiomas, sino que, para capacidades especiales, ya que puede emplear diferentes técnicas de enseñanza y a través de la IAG adaptar el contenido de acuerdo con la capacidad de aprendizaje del alumno.

La IA se emplea en las diversas actividades diarias como ya hemos comentado, sin embargo, estamos en el camino para que la IA se más inteligente y pueda tomar acciones y crear agentes de IA, como por ejemplo tener una asistente virtual que realice nuestras citas con base al aprendizaje de nuestros hábitos y preferencias, hasta poder tener listo el reporte para alguna reunión. Parece ficción, pero es parte de la evolución de tareas posibles desde IA predictiva, IA generativa y agentes de IA.

La IAG puede cambiar la vida de las personas, ya que puede aprender la forma en que nos comunicamos. Por ejemplo, una persona que desconoce el lenguaje de señas puede comunicarse en forma fluida con una persona muda

que se comunica mediante este lenguaje y traducir en voz las frases expresadas con señas. De igual manera, las personas con alguna limitación pueden aprender mediante un aplicativo de IA que se comunique y use un método mejor adaptado para la persona.

La conducción autónoma ya es una realidad. Los vehículos pueden llevarnos de un destino a otro, incluyendo sillas de ruedas para discapacitados y pueden aprender cuales son nuestras rutas de preferencia o simplemente tomar la más rápida, respetando las reglas de tránsito y normas de buena convivencia. Sin embargo, hay una gran cantidad de dilemas que resolver y que recaen en la regulación de cada país, pero esto no es un reto para la IA que puede aprender del reglamento y cultura vial del lugar en que se use.

En el ámbito tecnológico, hay una creciente necesidad para que se realice al borde en todas aquellas funciones que requieren de una respuesta inmediata y que no dependan de una red de comunicación para comunicarse a un servidor en la nube para ser procesadas. No solo en las funciones críticas y de emergencia, sino también para responder a las actividades de personas con capacidades especiales que se pueden encontrar en un área sin cobertura de comunicación.

El mayor desafío tecnológico para desarrollar dispositivos con IA al borde radica principalmente en aquellos de uso personal, para lograr una buena experiencia que requiere de un tamaño pequeño, ligero y ergonómico, que sea fácil de transportar a todas partes, así como una autonomía de varias horas o días con una sola carga de batería, que tampoco debe ser ni grande ni pesada. Por ello son imprescindibles hacer inversiones de varios billones de dólares en investigación y desarrollo de semiconductores de menor tamaño de nodo que cada día desafían las leyes de la física, al tiempo que se diseñan nuevos algoritmos más eficientes y optimizados, que realicen el autoaprendizaje más rápido y preciso, así como técnicas innovadoras para comprimir el manejo de memoria. Como mencionamos anteriormente los sistemas en chip semiconductor son una respuesta, ya que no solo han logrado integrar procesadores de cómputo, gráficos y de inteligencia artificial cada vez más potentes, sino también procesadores de imagen de cámara, controladores de pantallas digitales, manejo de sensores y periféricos, decodificadores de audio y video, así como de módems de comunicación inalámbrica tanto móvil como satelital, Wi-Fi y bluetooth, sistemas de posicionamiento satelital global. Todo esto y más en un diminuto encapsulado.

DESAFÍOS POLÍTICOS Y REGULATORIOS

Mientras la tecnología avanza a pasos agigantados, los gobiernos enfrentan un gran reto para regular las aplicaciones de la IA que responda a los requerimientos de los diferentes grupos sociales, con una visión ética, con respeto a los derechos humanos y de la naturaleza en general.

En resumen, la regulación de la IA debe enfocarse al mejoramiento de la vida de las especies en todos sus aspectos y cuidado de su entorno natural.

La comunidad europea ha hecho un gran avance sobre las políticas y regulación de la IA, y cada país debe definir sus propias políticas públicas y estrategia nacional, para regular con ética y con pleno respeto de los derechos humanos. Por lo que la colaboración entre las diferentes naciones ayudará a definir regulaciones universalmente aceptadas, así como también es un buen camino para que los países menos desarrollados logren adoptar y/o adaptar las mejores prácticas de las políticas públicas y regulaciones, a través del aprendizaje y experiencia de los países que ya transitaron por este camino.

V. REFLEXIONES FINALES

TENDENCIAS EMERGENTES EN IA

La inteligencia artificial (IA) está evolucionando rápidamente, y con ello surgen nuevas tendencias que están transformando diferentes sectores. Algunas de las tendencias emergentes más destacadas en IA incluyen:

IA EN LA MEDICINA PERSONALIZADA

La IA está avanzando en el campo de la medicina, particularmente en la creación de tratamientos personalizados basados en el análisis de grandes volúmenes de datos. La IA puede analizar registros médicos, datos genéticos y otros factores para:

- Desarrollar terapias a medida.
- Identificar nuevos biomarcadores para enfermedades.
- Mejorar el diagnóstico precoz y la predicción de resultados.

IA en la Automatización de Procesos Empresariales (RPA)

La automatización de procesos mediante IA (Robotic Process Automation, RPA) está ayudando a las empresas a reducir costos y mejorar la eficiencia. Los sistemas de IA combinados con RPA son capaces de realizar tareas repetitivas y complejas, como:

- Procesar documentos y formularios.
- Gestionar bases de datos.
- Interactuar con sistemas de clientes y proveedores.

Inteligencia Artificial Multimodal

Los modelos multimodales son capaces de integrar y procesar información de diferentes tipos de datos (texto, imagen, audio, etc.) de manera conjunta. Esto permite que la IA pueda:

- Comprender y generar contenido en múltiples formatos (por ejemplo, modelos que pueden generar imágenes a partir de descripciones textuales o viceversa).
- Interacción más natural con asistentes virtuales, que pueden entender y responder no solo a comandos de texto, sino también a imágenes, sonidos y gestos.

IA Y CIBERSEGURIDAD

La IA está jugando un papel crucial en la ciberseguridad, ayudando a detectar patrones y predecir posibles amenazas de forma proactiva. Algunos ejemplos incluyen:

- Detección de anomalías en redes y sistemas.
- Análisis predictivo para identificar vulnerabilidades antes de que sean explotadas.
- Autodefensa mediante sistemas que pueden reaccionar a ataques en tiempo real.

TRANSFORMACIÓN DE LA FUERZA LABORAL (IA EN RECURSOS HUMANOS)

Las tecnologías de IA están revolucionando la gestión de recursos humanos, desde la contratación hasta el análisis de desempeño. Algunas aplicaciones incluyen:

- **Reclutamiento basado en IA:** Análisis de currículums, entrevistas automatizadas y recomendaciones de candidatos.
- **Análisis de bienestar de empleados:** Sistemas que analizan el comportamiento para predecir la rotación o el agotamiento de los empleados.
- **Desarrollo de talento:** Plataformas que proporcionan capacitación personalizada a los empleados.

IA EN EL EDGE COMPUTING

La computación en el borde (edge computing) está ganando popularidad al combinar la IA con dispositivos locales (como sensores, cámaras o dispositivos IoT). Esto permite ejecutar modelos de IA cerca de la fuente de datos, lo que reduce la latencia y mejora la eficiencia:

- Aplicaciones en tiempo real como vehículos autónomos, sistemas de vigilancia o diagnósticos médicos instantáneos.
- Menor dependencia de la nube, lo que puede ser crucial en entornos con conexión limitada.

Modelos de IA Autónomos y Robótica

La robótica y los vehículos autónomos continúan avanzando con la ayuda de la IA. Esto abarca desde robots industriales hasta vehículos autónomos y drones. Los modelos autónomos están mejorando en áreas como:

- Transporte autónomo: Vehículos sin conductor y drones de entrega.
- Robots colaborativos (cobots) en fábricas y entornos laborales.
- Automatización en entornos no industriales como la asistencia doméstica o la atención a personas mayores.

Redes Neuronales de Última Generación y Arquitecturas Híbridas

Las redes neuronales están evolucionando con arquitecturas cada vez más sofisticadas. Algunas de las innovaciones incluyen:

- Redes de atención (como Transformers) que han mejorado el procesamiento de lenguaje natural y visión computacional.
- Arquitecturas híbridas que combinan modelos de IA simbólica con aprendizaje profundo, con el fin de mejorar la razonabilidad y la generalización de las máquinas.

IA en la Sostenibilidad y el Cambio Climático

La IA también está ayudando en la lucha contra el cambio climático y la sostenibilidad mediante:

- Modelos predictivos para evaluar los efectos del cambio climático.
- Optimización energética: Sistemas de IA que ajustan el uso de energía en edificios y fábricas para reducir emisiones.
- Agricultura inteligente: Aplicación de IA para optimizar el uso de recursos en la agricultura y mejorar la producción sostenible de alimentos.

IA para el Diseño y la Creación de Videojuegos

La IA está ganando terreno en el desarrollo de videojuegos, desde el diseño de niveles hasta la creación de personajes no jugables (NPCs) más complejos. Algunos avances incluyen:

- Generación procedural de mundos y niveles.
- Comportamientos adaptativos de NPCs, que reaccionan de manera más realista a las acciones del jugador.
- Mejora de la jugabilidad con IA que ajusta la dificultad según las habilidades del jugador.

A manera de resumen

Estas tendencias emergentes no solo están transformando la forma en que interactuamos con la tecnología, sino que también están teniendo un impacto significativo en industrias clave, como la salud, la manufactura, el transporte y la educación. A medida que la IA continúa evolucionando, veremos más avances que desafían los límites de lo que es posible, creando nuevas oportunidades y también desafíos para las sociedades y las economías globales.

Proyecciones y Escenarios Futuros

La inteligencia artificial (IA) está evolucionando rápidamente y se espera que tenga un impacto profundo en diversas áreas en el futuro. Las proyecciones y los escenarios futuros de la IA varían dependiendo de las áreas específicas en las que se enfoque, pero en términos generales se pueden identificar algunos temas clave y tendencias que probablemente marcarán el desarrollo y la integración de la IA en la sociedad y la economía global. A continuación, se presentan algunas proyecciones y escenarios futuros sobre la IA:

IA en la automatización y el empleo

- Automatización de trabajos repetitivos y peligrosos: Se espera que la IA continúe mejorando la automatización en sectores como la manufactura, la logística, y los servicios. Esto implicará una mayor eficiencia y reducción de costos en diversas industrias, pero también podría generar dislocaciones en el empleo, especialmente en trabajos que dependen de habilidades repetitivas o físicas.
- Recalificación laboral y nuevas oportunidades: Con la automatización de ciertos trabajos, es probable que emerjan nuevas oportunidades laborales en áreas relacionadas con la programación, gestión de IA, análisis de datos y mantenimiento de sistemas automatizados. Sin embargo, también será

necesario un esfuerzo significativo en la capacitación y recalificación de los trabajadores.

- Transformación de la educación: A medida que la IA avance, se espera que el sistema educativo evolucione para formar a los estudiantes en habilidades que sean complementarias a la IA, como el pensamiento crítico, la creatividad, y las habilidades sociales que las máquinas no pueden replicar fácilmente.

IA en la medicina y la salud

- Diagnóstico y personalización del tratamiento: La IA tiene el potencial de transformar la atención médica, desde la mejora en los diagnósticos hasta la creación de tratamientos más personalizados basados en los datos genéticos de los pacientes. El uso de IA para analizar grandes cantidades de datos médicos permitirá identificar patrones y prever enfermedades con mayor precisión.
- Asistentes virtuales en salud: Con la mejora de la IA en el procesamiento del lenguaje natural, los asistentes virtuales (como los chatbots de salud) serán más efectivos en el diagnóstico y la recomendación de tratamientos médicos. También podrían ofrecer apoyo en la gestión de enfermedades crónicas y el monitoreo de la salud en tiempo real.
- IA en la investigación de fármacos: Las plataformas basadas en IA podrían acelerar el proceso de descubrimiento de fármacos al analizar vastas bases de datos de compuestos químicos y simulaciones moleculares. Esto podría reducir el tiempo y los costos necesarios para desarrollar nuevos medicamentos.

IA en el ámbito de la ética y la gobernanza

- Desarrollo de marcos éticos para la IA: A medida que la IA se integra más en la sociedad, se incrementarán los debates sobre su impacto ético y social. Habrá una creciente necesidad de marcos regulatorios internacionales que garanticen que la IA se desarrolle de manera justa, transparente y responsable. Esto incluye temas como la privacidad, la toma de decisiones autónoma y la mitigación de sesgos en los algoritmos.
- IA en la toma de decisiones políticas: Es probable que se utilicen algoritmos de IA para ayudar en la toma de decisiones políticas y gubernamentales, analizando grandes volúmenes de datos y proporcionando recomendaciones para resolver problemas sociales y económicos. Sin embargo, esto también podría generar preocupaciones sobre el control y la transparencia en estos sistemas.

IA en la economía y la productividad

- Aumento de la productividad: En muchos sectores, la IA contribuirá a un aumento general de la productividad. Desde la optimización de cadenas de suministro hasta la mejora de la eficiencia en la industria financiera, la IA facilitará un mejor uso de los recursos y permitirá realizar tareas de forma más rápida y precisa.
- IA y el crecimiento de las "economías de plataformas": Las plataformas basadas en IA (por ejemplo, aplicaciones de transporte, comercio electrónico y servicios financieros) continuarán proliferando, cambiando la forma en que las empresas y los consumidores interactúan. Las empresas tendrán acceso a tecnologías avanzadas sin necesidad de tener infraestructura propia, lo que democratiza el acceso a soluciones tecnológicas de vanguardia.
- IA y la distribución de la riqueza: Un desafío importante será cómo la IA afectará la distribución de la riqueza. La automatización de trabajos y el uso de la IA por parte de grandes corporaciones podrían exacerbar las desigualdades económicas. Algunos expertos sugieren que se necesitan nuevos enfoques de políticas públicas, como impuestos sobre la automatización o la implementación de un ingreso básico universal, para contrarrestar estos efectos.

IA en la sociedad y la cultura

- Transformación de las interacciones sociales: A medida que los sistemas de IA mejoren, cambiarán la forma en que las personas interactúan entre sí y con las máquinas. Los asistentes personales basados en IA (como los de voz y los chatbots) se integrarán aún más en la vida cotidiana, cambiando la naturaleza de la comunicación, el entretenimiento y el consumo de información.
- IA generativa en arte y creatividad: La IA generativa, como los modelos de lenguaje y los generadores de imágenes (por ejemplo, los modelos como GPT y DALL·E), revolucionará la creación de contenido. Esto podría cambiar la industria del entretenimiento, la publicidad y las artes, al permitir la creación rápida y eficiente de contenido visual, textual y audiovisual. Sin embargo, también podrían surgir preguntas sobre los derechos de autor y la autenticidad.
- IA y el comportamiento humano: En un mundo donde los algoritmos de IA tienen un papel cada vez más importante en la toma de decisiones, se podría producir una influencia cada vez mayor en el comportamiento humano, desde las decisiones de compra hasta las preferencias políticas. Las plataformas de

redes sociales ya utilizan IA para personalizar la experiencia del usuario, y se espera que este enfoque se expanda.

Desafíos y riesgos futuros de la IA

- Riesgos de la inteligencia artificial avanzada: A medida que los sistemas de IA se vuelven más sofisticados, algunos investigadores advierten sobre los posibles riesgos existenciales que podrían derivarse de la creación de inteligencia artificial general (AGI). Estos riesgos incluyen la pérdida de control sobre sistemas autónomos y la posibilidad de que la IA desarrolle metas incompatibles con los intereses humanos.
- IA y la ciberseguridad: Los sistemas de IA también pueden ser utilizados para mejorar la ciberseguridad, pero también plantean nuevas vulnerabilidades. Los atacantes podrían utilizar IA para llevar a cabo ciberataques más sofisticados y automatizados, lo que requeriría una defensa igualmente avanzada.
- El control social y la privacidad: Con la recopilación masiva de datos y la capacidad de la IA para analizar patrones de comportamiento, surgen preocupaciones sobre la invasión de la privacidad y el uso indebido de la información personal. La vigilancia automatizada y la toma de decisiones autónomas también pueden generar tensiones sobre el control social y la libertad individual.

La IA del Futuro

El futuro de la inteligencia artificial es prometedor pero incierto. Mientras que la IA tiene el potencial de transformar muchas áreas de la vida humana de manera positiva, también presenta desafíos significativos que deben ser gestionados cuidadosamente. La clave será asegurarse de que su desarrollo sea inclusivo, ético y orientado a beneficios para la humanidad en su conjunto, a la vez que se gestionan los riesgos asociados con su adopción a gran escala. El equilibrio entre la innovación, la regulación y la ética será fundamental para un futuro en el que la IA desempeñe un papel central en la sociedad.

Estrategias para Superar los Desafíos

Superar los desafíos de la inteligencia artificial (IA) es una tarea compleja que involucra aspectos técnicos, éticos, sociales y económicos. Algunas estrategias clave para enfrentarlos, tanto a nivel individual como organizacional, incluyen:

Fomentar la Educación y Capacitación Continua

- Capacitación de habilidades técnicas: La IA está en constante evolución. Para enfrentarse a los desafíos que presenta, es esencial que los profesionales estén al tanto de los últimos avances y desarrollos en algoritmos, herramientas y metodologías de IA.
- Formación ética: La educación no debe centrarse solo en la parte técnica, sino también en la ética de la IA. Los trabajadores deben aprender a tomar decisiones informadas sobre cómo usar la IA de manera responsable y justa.
- Programas educativos inclusivos: Crear programas que sean accesibles para todos, desde estudiantes hasta profesionales en la industria, para evitar la brecha de habilidades en la fuerza laboral.

Desarrollar Marcos Regulatorios y Éticos

- Políticas públicas claras: Los gobiernos deben establecer regulaciones claras que guíen el desarrollo y la implementación de IA, asegurando la privacidad, la seguridad y la justicia social. Ejemplos de marcos regulatorios incluyen las leyes de protección de datos como el GDPR en Europa.
- Ética en el diseño y uso: Asegurar que la IA se desarrolle con principios éticos en mente, tales como la equidad, la transparencia, la rendición de cuentas y la no discriminación.
- Colaboración entre sectores: Gobiernos, empresas y organizaciones civiles deben colaborar para crear políticas públicas que favorezcan el desarrollo seguro y beneficioso de la IA.

Enfrentar los Desafíos del Empleo y la Automatización

- Reentrenamiento y re-skilling: A medida que los trabajos cambian debido a la automatización, las estrategias de reentrenamiento son cruciales. Las políticas laborales deben incluir programas que ayuden a los trabajadores a adaptarse a los nuevos roles.
- Creación de nuevos empleos: A medida que algunas tareas sean automatizadas, otras nuevas oportunidades laborales emergen. Es importante fomentar sectores emergentes como la IA, el análisis de datos, la ciberseguridad, la biotecnología y otros campos que requieren habilidades especializadas.
- Adaptación a la transición: Es necesario facilitar una transición justa para los trabajadores que podrían verse desplazados por la automatización. Esto incluye apoyo económico, programas de capacitación y políticas públicas que ofrezcan seguridad laboral.

Mejorar la Transparencia y Explicabilidad de los Modelos de IA

- Explicabilidad de los algoritmos: Es importante que las decisiones tomadas por sistemas de IA sean comprensibles para los humanos. Implementar métodos de IA explicables (XAI, por sus siglas en inglés) ayuda a garantizar que las decisiones de la IA sean auditables y comprensibles.
- Transparencia en el uso de datos: Las organizaciones deben ser claras sobre qué datos se utilizan y cómo se recopilan, procesan y almacenan. Los usuarios deben tener control sobre sus propios datos.
- Auditorías independientes: Desarrollar mecanismos que permitan auditar los algoritmos y garantizar que no estén sesgados, discriminatorios ni violen normas éticas.

Garantizar la Seguridad y Privacidad

- Ciberseguridad robusta: Con el aumento de la integración de la IA en sistemas críticos, garantizar la seguridad cibernética se vuelve más importante que nunca. La IA puede ser vulnerable a ataques que exploten sus debilidades, por lo que es esencial invertir en protección avanzada.
- Protección de la privacidad: La IA a menudo requiere grandes cantidades de datos, muchos de los cuales son personales. Las estrategias para manejar y proteger esos datos de manera segura deben ser una prioridad para evitar brechas de seguridad o el uso indebido de la información.

Promover la Colaboración Interdisciplinaria

- Trabajo conjunto entre humanos y máquinas: La IA debe verse como una herramienta que complementa las capacidades humanas en lugar de sustituirlas completamente. Fomentar la colaboración entre máquinas y humanos puede potenciar las habilidades de ambos.
- Integración de diversas disciplinas: Los avances en IA no solo requieren conocimientos en informática, sino también en áreas como psicología, sociología, filosofía, derecho, economía y más. Una colaboración interdisciplinaria permitirá abordar los problemas de la IA desde diferentes perspectivas.

Reducir el Sesgo y la Discriminación en la IA

- Diversidad en los equipos de desarrollo: Incluir equipos diversos en el desarrollo de la IA, con diferentes antecedentes, experiencias y perspectivas, puede ayudar a reducir el sesgo en los modelos.

- Auditorías y pruebas para el sesgo: Las organizaciones deben realizar pruebas rigurosas para detectar sesgos en sus sistemas de IA, especialmente en áreas como la contratación, los préstamos y la justicia penal.

Fomentar la Responsabilidad Social y la Conciencia Pública

- Conciencia pública sobre IA: Es fundamental que el público general entienda qué es la IA, cómo afecta sus vidas y qué riesgos y beneficios conlleva. Programas educativos, charlas públicas y divulgación científica pueden ayudar en este aspecto.
- Participación ciudadana en decisiones tecnológicas: Los ciudadanos deben tener voz en la forma en que la IA se utiliza en sus sociedades. La democracia y la participación son clave para asegurar que la IA se desarrolle en beneficio de la humanidad.

Desarrollar y Adoptar Soluciones Escalables

- Implementación escalable: A medida que las soluciones basadas en IA se implementan a gran escala, las organizaciones deben asegurarse de que sean escalables y sostenibles, tanto desde el punto de vista tecnológico como económico.
- Innovación abierta: Las iniciativas de innovación abierta y colaboración entre empresas, gobiernos y universidades pueden acelerar el desarrollo de soluciones de IA que sean beneficiosas para la sociedad en su conjunto.

Prepararse para los Desafíos Futuros

- Anticipar problemas éticos y sociales: La IA plantea desafíos complejos, como la creación de superinteligencias o la posible manipulación de la información. Las instituciones deben ser proactivas en la identificación de estos problemas y en la preparación de soluciones.
- Preparación ante la disrupción tecnológica: Aunque la IA traerá enormes beneficios, también puede causar disrupciones significativas en diferentes industrias y en la sociedad. Es importante tener planes de contingencia para manejar estos cambios.

A manera de resumen

Superar los desafíos que presenta la inteligencia artificial requiere un enfoque multifacético y colaborativo. Desde el fortalecimiento de la educación y la ética hasta la creación de marcos regulatorios, cada acción es un paso hacia un

futuro en el que la IA se desarrolle de manera beneficiosa y equitativa para la humanidad.

Visión a Largo Plazo para la Región

La visión a largo plazo para la región de México y América Latina en el ámbito de la inteligencia artificial (IA) puede tomar diferentes formas, pero en términos generales, la clave está en convertir a la región en un actor relevante en el ecosistema global de la IA, aprovechando su potencial humano, recursos y su diversidad cultural. Aquí se plantean algunas direcciones clave para esta visión:

Desarrollo de Talento Local y Capacitación

Una de las piedras angulares de la visión a largo plazo es la construcción de una infraestructura educativa robusta y accesible que forme a miles de profesionales en IA. Esto implica tanto la capacitación de científicos de datos e ingenieros como el fomento de habilidades interdisciplinares que combinen conocimiento en IA con áreas clave como la ética, la economía, la biología, la salud pública, etc.

Acciones concretas:

- Creación de programas académicos especializados en IA a nivel universitario.
- Inversión en plataformas de aprendizaje accesibles y de bajo costo, como cursos en línea y bootcamps.
- Colaboración entre universidades, gobiernos y empresas privadas para crear centros de investigación y de innovación en IA.

Investigación y Desarrollo (I+D) Regional

A largo plazo, América Latina debe apostar por ser un referente global en investigación en áreas de IA que sean de relevancia tanto para los problemas locales como para el contexto global. Áreas como la inteligencia artificial aplicada al cambio climático, la salud, la agricultura, la energía renovable y la mejora de procesos en sectores industriales serán claves.

Acciones concretas:

- Incentivar la creación de centros de excelencia en investigación de IA, financiados tanto por fondos gubernamentales como por inversiones privadas.
- Establecer alianzas internacionales de investigación para que las universidades y centros de I+D latinoamericanos puedan participar en redes de colaboración globales.

Infraestructura Tecnológica y Ecosistema de Innovación

El crecimiento de la IA está profundamente ligado a la infraestructura tecnológica. América Latina necesita mejorar la infraestructura digital y tecnológica, promoviendo una conectividad de alta velocidad en áreas urbanas y rurales, el desarrollo de infraestructuras de nube locales y la adopción de tecnologías emergentes.

Acciones concretas:

- Promover políticas públicas para el acceso universal a internet de alta velocidad, que permita la conectividad en áreas rurales y periféricas.
- Incentivar la creación de hubs tecnológicos, como parques tecnológicos y espacios colaborativos, en los que las startups tecnológicas puedan desarrollarse, crecer y colaborar entre sí.
- Mejorar el acceso a plataformas en la nube y potenciar la adopción de tecnologías como el 5G para acelerar la digitalización.

Adopción de la IA en Sectores Estratégicos

Para que la IA sea un motor de desarrollo, debe ser adoptada en sectores estratégicos que puedan transformar profundamente la economía y la sociedad. En América Latina, sectores como la agricultura, la salud, la educación, la energía y las ciudades inteligentes tienen un gran potencial para aplicar soluciones basadas en IA.

Acciones concretas:

- Fomentar la adopción de IA en la agricultura para mejorar la productividad y sostenibilidad mediante el uso de tecnologías como drones, sensores y análisis de datos para la gestión de cultivos y recursos.
- Desarrollar soluciones de IA en salud, tanto en diagnóstico como en la administración pública, para mejorar la atención a pacientes y la eficiencia de los sistemas de salud pública.
- Incentivar el uso de IA para la optimización de la cadena de suministro y la automatización de procesos industriales.

Ética y Regulación de la IA

La IA plantea desafíos éticos y sociales importantes, por lo que será fundamental que los países latinoamericanos desarrollen marcos regulatorios que promuevan el desarrollo responsable y ético de estas tecnologías. Esto incluiría garantizar la privacidad de los datos, prevenir sesgos algorítmicos y asegurarse de que las tecnologías sean inclusivas y no excluyan a poblaciones vulnerables.

Acciones concretas:

- Desarrollar marcos regulatorios nacionales y regionales que promuevan la ética en la IA, asegurando la transparencia, la justicia y la privacidad.
- Fomentar la creación de comités y organizaciones dedicadas a la investigación de los impactos sociales y económicos de la IA, con un enfoque en la protección de los derechos humanos.
- Crear alianzas internacionales para colaborar en estándares globales para la IA ética, permitiendo que América Latina influencie las políticas globales en este campo.

Inclusión y Diversidad en el Desarrollo de IA

América Latina es una región diversa en términos de cultura, etnias y lenguas, lo que ofrece una gran oportunidad para crear soluciones de IA que consideren esta diversidad. La visión a largo plazo debe incluir la creación de tecnologías que sean inclusivas y accesibles para todos, incluyendo las comunidades indígenas y rurales, y considerando la pluralidad de idiomas y contextos sociales.

Acciones concretas:

- Desarrollar modelos de IA multilingües que reconozcan los diferentes idiomas y dialectos de la región, como el náhuatl, quechua, guaraní, entre otros.
- Fomentar la participación activa de comunidades marginadas en el diseño y desarrollo de tecnologías, asegurando que la IA sirva para cerrar brechas sociales en lugar de ampliarlas.

Colaboración Regional y Global

A largo plazo, América Latina debe promover una visión colectiva sobre el desarrollo de la IA. Esto implica la cooperación entre gobiernos, empresas y academia dentro de la región, así como la integración en redes y proyectos globales de IA.

Acciones concretas:

- Fomentar acuerdos y asociaciones estratégicas entre países latinoamericanos para compartir conocimiento, experiencia y recursos.
- Participar activamente en foros internacionales sobre IA, estableciendo alianzas con potencias tecnológicas como Estados Unidos, la Unión Europea y China, pero también con otros países emergentes en IA.

Desarrollo Sostenible y Responsabilidad Social

La IA puede jugar un papel fundamental en la solución de problemas globales como el cambio climático, la pobreza y la inseguridad alimentaria. La región

debe enfocarse en una visión de desarrollo sostenible donde la inteligencia artificial sea una herramienta para la inclusión social y económica, y no para perpetuar desigualdades.

Acciones concretas:

- Desarrollar y aplicar soluciones basadas en IA para la sostenibilidad, como la gestión eficiente de recursos naturales y la reducción de la huella de carbono en las industrias.
- Asegurar que las políticas y aplicaciones de IA favorezcan la igualdad de oportunidades, la equidad de género y la inclusión social.

A manera de resumen

La visión a largo plazo para la región de América Latina en inteligencia artificial debe ser inclusiva, ética, sostenible e innovadora. Aprovechando el talento humano, la infraestructura y los recursos naturales, y promoviendo la cooperación entre países y actores internacionales, América Latina puede convertirse en un centro global de innovación en inteligencia artificial, contribuyendo significativamente a los avances globales en esta área mientras aborda los desafíos locales. Para ello, será crucial la colaboración entre el sector público, privado, la academia y la sociedad civil, asegurando que la IA sea una herramienta para el desarrollo equitativo y sostenible.

VI. ANEXO: APÉNDICES

A. Columna de Opinión: ¿Es la IA Generativa únicamente para grandes empresas?
B. Infográfico: Guía de innovación de IoT: AI y el futuro de la fabricación industrial
C. Infográfico: ¿Qué es la IA Generativa y por qué es importante?
D. Blog: Cómo acelerar Wi-Fi con IA
E. Blog: La nueva ola de innovación automotriz basada en Inteligencia Artificial
F. Blog: Seguridad y eficiencia energética en el corazón de la Edge AI
G. Video: Democratizando las técnicas de computación de alto rendimiento
H. Video: El camino de MediaTek como líder en el mercado móvil
I. Video: Redefiniendo el futuro con la investigación avanzada de MediaTek
J. Video: MediaTek Dimensity 9400 – Poder hasta el Núcleo
K. Video: La IA en el dispositivo de MediaTek crea nuevas posibilidades en móviles
L. Video: La diversificación hace a MediaTek #1
M. Video: Una nueva era móvil con el MediaTek Dimensity 9400
N. Video: MediaTek Dimensity 9400: GPU G925 con Raytracing y Opacity Micromaps (OMM)
O. Video: Destaca con MediaTek: rendimiento y tecnología a la vanguardia
P. Video: MediaTek Dimensity 9400: 1er generador de video con Gen-IA de alta calidad impulsado por smartphones
Q. Video: Colaboración Google x MediaTek - Dimensity 9400 y Google Gemini Nano con multimodalidad

R. Video: Wi-Fi 7 y el avance de MLO: transformando la conectividad
S. Video: Estableciendo la base en MediaTek
T. Video: 5G: Mantente siempre conectado con MediaTek
U. Video: Transformando las experiencias de juego de nueva generación con MediaTek
V. Video: Descubre la magia de los teléfonos inteligentes con MediaTek
W. Video: Explorando la IA similar a la humana con MediaTek
X. Video: MediaTek está desplegando nuevos beneficios del 5G a través de RedCap
Y. Video: Wi-Fi 7: Mejores velocidades y rendimiento con MediaTek
Z. Video: Revolucionando los vehículos con IA: innovación MediaTek
AA. Video: Democratizando las técnicas de computación de alto rendimiento con MediaTek
BB. Video: Hogar inteligente y más allá con MediaTek
CC. Video: Tendencias de la industria de los smartphones con MediaTek
DD. Video: Aprovechando la IA Generativa en el sector privado
EE. Video: Punto de Vista MediaTek: Las tendencias a tener en cuenta

VII. BIBLIOGRAFÍA

CEPAL (Comisión Económica para América Latina y el Caribe) (2024a), Panorama de las Políticas de Desarrollo Productivo en América Latina y el Caribe, 2024 (LC/PUB.2024/15-P), Santiago. (2024b), América Latina y el Caribe ante las trampas del desarrollo: transformaciones indispensables y cómo gestionarlas (LC/SES.40/3-P),

https://www.cepal.org/es/publicaciones/80841-superar-trampas-desarrollo-america- latina-caribe-la-era-digital-potencial

https://repositorio.cepal.org/server/api/core/bitstreams/f096da2a-5107-486c-94f3- 871683423556/content

Santiago. (2024c), CEPALSTAT [base de datos en línea] https://statistics.cepal.org/portal/cepalstat/index.html?lang=es

CAPÍTULO TRES

INTELIGENCIA ARTIFICIAL. LAS NUEVAS POSIBILIDADES DEL DESARROLLO ECONÓMICO

FUTUROS IMAGINADOS, EXPECTATIVAS FICCIONALES Y DINÁMICA DEL CAPITALISMO; EL PAPEL DE LA INTELIGENCIA ARTIFICIAL

SUMARIO

I. Futuros imaginados y expectativas ficcionales II. Impactos sociales y éticos **III.** Posibilidades para un capitalismo diferente **IV.** Dinámica del capitalismo en la era de la IA **V.** Conclusión y reflexiones finales **VI.** Bibliografía

> CLEMENTE RUIZ DURÁN

Profesor e investigador del Posgrado de Economía de la Facultad de Economía de la UNAM. La elaboración de este documento contó con el apoyo de Miguel Ángel Hernández Ruiz, Rafael Ulises Ramírez López y Jean Pablo Vite Sixto

La interacción entre futuros imaginados, expectativas ficcionales y las dinámicas del capitalismo adquiere una nueva relevancia en la era de la inteligencia artificial (IA). En este documento se propone explorar el tema a partir de algunas reflexiones claves que permitan construir escenarios que estructuren la realidad que estamos construyendo a partir de la irrupción de la inteligencia artificial.

I. FUTUROS IMAGINADOS Y EXPECTATIVAS FICCIONALES

El concepto de futuros imaginados tal como lo conceptualizó Jens Beckert en su libro "Fictional Expectations and Capitalist Dynamics: Imagined Futures" se relaciona con cómo las sociedades anticipan el devenir a través de narrativas, mitos y representaciones culturales. En el contexto del capitalismo, estas expectativas ficcionales funcionan como un motor que impulsa decisiones económicas, políticas y tecnológicas. En la actualidad el auge de las "ciudades inteligentes" y la expectativa de automatización total se basan más en visiones ficcionales que en realidades concretas. Narrativas como la "singularidad tecnológica" o la colonización de Marte reflejan aspiraciones capitalistas que moldean la inversión y el desarrollo científico-tecnológico. El capitalismo se alimenta de estas ficciones para perpetuar el ciclo de producción y consumo. Las promesas de innovación y progreso se convierten en commodities que venden futuros específicos, a menudo favoreciendo a corporaciones tecnológicas.

El capitalismo, caracterizado por su capacidad para absorber y transformar cualquier innovación en valor económico, encuentra en la IA un aliado estratégico, esto se manifiesta en la aceleración del mercado al optimizar procesos de producción, distribución y consumo, lo que fortalece aún más el

dominio de gigantes digitales como Amazon, Google y Meta. Esta transformación se basa en la explotación masiva de datos personales, considerados el nuevo "petróleo", con la IA como herramienta clave para extraerles valor, lo que amplifica las desigualdades existentes. Además, muchas empresas de IA promueven la idea de que su tecnología está destinada a "disrumpir" industrias enteras, alimentando un ciclo interminable de innovación y obsolescencia programada que pone en duda la estabilidad a largo plazo de las estructuras económicas y sociales.

La IA como creadora y destructora de ficciones, tenemos que estar conscientes de que la inteligencia artificial no solo ejecuta tareas, sino que participa activamente en la creación de futuros ficcionales: Los algoritmos generativos, como los modelos de lenguaje, tienen la capacidad de crear relatos futuristas, campañas publicitarias y hasta predicciones económicas que pueden influir en el imaginario colectivo. Sin embargo, esta tecnología también plantea desafíos importantes, como la erosión de la autenticidad. La proliferación de deepfakes y narrativas creadas por IA pone en duda la confianza en las fuentes tradicionales de información, dificultando la veracidad de lo que consumimos. Así como el creciente debate de la energía que será necesaria para el aumento de la demanda por esta tecnología. Además, las decisiones automatizadas basadas en predicciones algorítmicas están teniendo un impacto tangible en áreas como los mercados financieros, las políticas públicas y el comportamiento social, lo que resalta el poder creciente que esta teniendo la IA en nuestra vida cotidiana.

II. IMPACTOS SOCIALES Y ÉTICOS

La intersección entre expectativas ficcionales y dinámicas del capitalismo exacerbadas por la IA plantea desafíos éticos y sociales:

La exclusión digital es una consecuencia significativa del avance de la inteligencia artificial y otras tecnologías avanzadas. A medida que estas tecnologías se integran cada vez más en diversos aspectos de la vida cotidiana, las personas que no tienen acceso a ellas o carecen de las habilidades necesarias para utilizarlas pueden quedar marginadas. Esto puede resultar en una brecha creciente entre aquellos que pueden beneficiarse de las oportunidades que ofrece la tecnología y aquellos que no, exacerbando desigualdades sociales y económicas. Es crucial abordar esta brecha digital para asegurar que todos tengan la oportunidad de participar en el futuro tecnológico.

La inteligencia artificial juega un papel crucial en el capitalismo de vigilancia, donde las tecnologías avanzadas se utilizan para monitorear a los espacios

públicos. Las promesas de la IA se entrelazan con sistemas de control social, como el reconocimiento facial y las plataformas de crédito social, que permiten un seguimiento y evaluación detallada de los individuos. El reconocimiento facial puede identificar y rastrear a personas en tiempo real en espacios públicos, mientras que las plataformas de crédito social pueden clasificar a los ciudadanos según su comportamiento y cumplimiento de normas, afectando su acceso a servicios y oportunidades. Aunque estas aplicaciones de la IA pueden ofrecer beneficios en términos de seguridad y eficiencia, también plantean serias preocupaciones sobre la privacidad, la libertad individual y el potencial de abuso de poder. Es fundamental desarrollar marcos éticos y legales que regulen el uso de estas tecnologías para proteger los derechos y libertades de las personas.

¿Qué futuros son posibles?: Los futuros imaginados suelen estar monopolizados por los medios, dejando de lado visiones alternativas de desarrollo más inclusivo y sostenible.

III. POSIBILIDADES PARA UN CAPITALISMO DIFERENTE

Una pregunta clave que surge de este replanteamiento de la realidad es la siguiente: ¿Es posible diseñar sistemas de IA que promuevan la equidad y la sostenibilidad? En este contexto de reimaginación de la realidad, los movimientos como el solarpunk o el decrecimiento ofrecen visiones alternativas para el futuro, proponiendo modelos que trascienden las dinámicas capitalistas tradicionales. Estos enfoques sugieren que es posible construir una sociedad más justa y sostenible, donde las tecnologías, incluidas las de inteligencia artificial, se utilicen para beneficio colectivo y no solo para unos pocos. En este sentido, si logramos garantizar que los beneficios de la IA estén al alcance de todos, podríamos mitigar las desigualdades que, de otro modo, se verían intensificadas por la implementación de estas tecnologías. De esta forma, la IA no solo podría contribuir a la eficiencia y la innovación, sino también convertirse en un motor de transformación social que impulse un futuro más equitativo y sostenible. *Esta interacción entre IA, narrativas del futuro y capitalismo sugiere que los "futuros imaginados" no son inevitables, sino que dependen de las decisiones políticas, éticas y sociales que tomemos hoy.* El surgimiento de esta nueva visión entrelaza elementos profundos de la sociología, la teoría económica y la filosofía de la tecnología. La interacción entre futuros imaginados, expectativas ficcionales y las dinámicas del capitalismo adquiere una nueva relevancia en la

era de la inteligencia artificial. A continuación, planteamos algunas reflexiones estructuradas para explorar este tema.

A. FUTUROS IMAGINADOS Y EXPECTATIVAS FICCIONALES.

Los futuros imaginados son proyecciones de la forma en que podría ser el mundo, edificadas a partir de sueños, anhelos y supuestos. A través de la historia, estos conceptos han sido un impulsor esencial para el progreso de la humanidad, dado que motivan políticas, proyectos de tecnología y estrategias de negocio. Ideas como las ciudades inteligentes o la colonización de Marte demuestran cómo estas perspectivas futuras pueden fomentar el progreso, aunque a menudo se fundamenten más en suposiciones que en situaciones que sean realmente tangibles.

En el capitalismo, estas narrativas no solo inspiran, sino que también se convierten en herramientas para movilizar recursos, atraer inversiones y justificar grandes proyectos. Las promesas de innovación y progreso se presentan como beneficios para todos, pero en la práctica, suelen concentrar los resultados en manos de unas pocas empresas o regiones, profundizando las desigualdades existentes. Después de todo, cada gran avance comienza como una simple idea.

Por eso, aunque estas visiones del futuro son necesarias para avanzar, debemos analizarlas críticamente. Solo así podremos asegurarnos de que las decisiones que tomemos hoy construyan un futuro más equitativo, sostenible e inclusivo, no solo para nuestra generación, sino para las que vienen. Resolver los problemas del presente con la mirada puesta en el bienestar de quienes nos seguirán es, en última instancia, lo que realmente importa.

IV. DINÁMICA DEL CAPITALISMO EN LA ERA DE LA IA

La inteligencia artificial ha transformado la forma en que operan los mercados, las industrias, las cadenas de valor y la producción. Uno de sus impactos más visibles es la aceleración del mercado, ya que optimiza procesos de producción, distribución y consumo, consolidando el poder de gigantes tecnológicos.

Las narrativas de disrupción se han convertido en un elemento central del discurso de muchas empresas de inteligencia artificial, que promocionan sus tecnologías como herramientas capaces de transformar radicalmente industrias enteras. Este enfoque no solo refuerza la idea de que el progreso tecnológico debe ser constante y acelerado, sino que también perpetúa dinámicas de

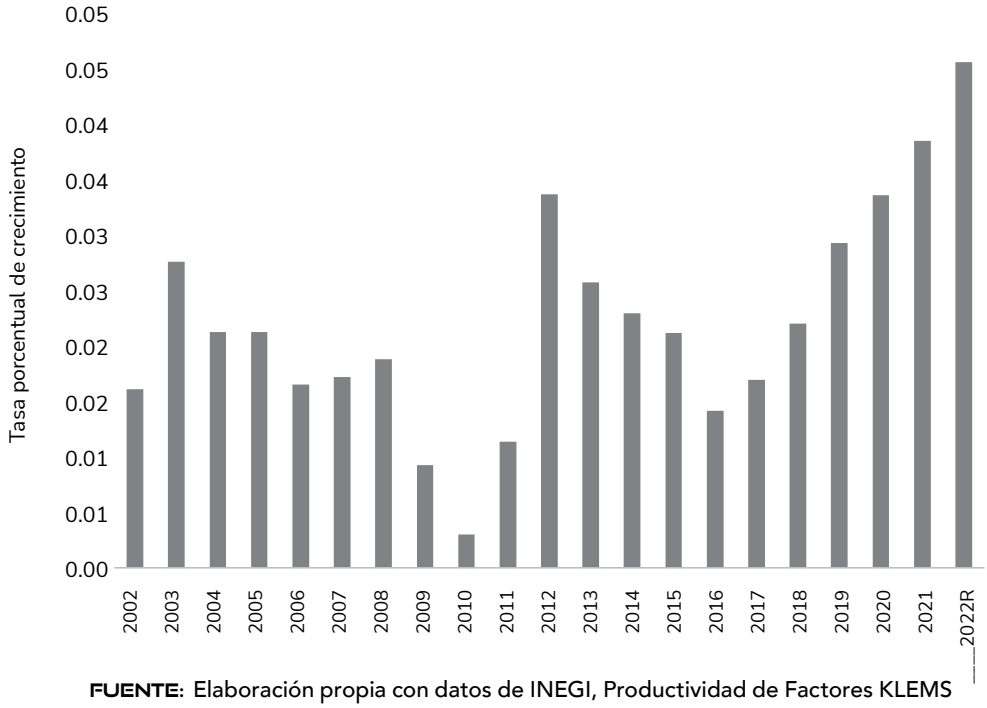
innovación continua que a menudo conducen a la obsolescencia programada. Al enfatizar la necesidad de adoptar las últimas herramientas y soluciones para no "quedar atrás", estas narrativas generan una presión constante sobre las empresas y los consumidores, incrementando el consumo de recursos y exacerbando desigualdades al privilegiar a quienes tienen acceso a las tecnologías más avanzadas. Además, esta visión de la innovación como un fin en sí mismo puede dificultar reflexiones más profundas sobre cómo la tecnología podría desarrollarse y aplicarse para promover equidad, sostenibilidad y bienestar colectivo.

En este escenario, el modelo de productividad total de los factores ampliado por el modelo KLEMS[1] se presenta como una herramienta valiosa para entender cómo sectores como el software en México pueden impulsar un crecimiento económico más inclusivo y sostenible. Este enfoque analiza de manera detallada, cómo factores como el capital, la mano de obra y los materiales contribuyen al desarrollo del país. Al conectar estos datos con las ideas de futuros alternativos, podemos imaginar cómo decisiones estratégicas podrían transformar la industria del software en una fuerza para reducir desigualdades y promover un equilibrio con el entorno. La gráfica a continuación nos ayuda a visualizar estas contribuciones y reflexionar sobre el papel que cada elección tiene en la construcción del futuro. La metodología KLEMS nos ayuda a entender de manera más clara cómo se compone el crecimiento económico, al desglosarlo en factores clave como el capital, el trabajo, la energía, los materiales y los servicios intermedios. En este caso, el gráfico 1 muestra cómo el sector del software ha contribuido al crecimiento económico de México entre 2002 y 2022, destacando el papel fundamental del capital intangible, como el software, en mejorar la productividad total de los factores (PTF).

A lo largo del período analizado, la contribución del software ha mostrado altibajos. En los años 2009 y 2011-2013 se observan picos significativos, probablemente impulsados por inversiones estratégicas en tecnología y procesos de transformación digital. Por otro lado, 2017 destaca por una notable caída, posiblemente relacionada con una reducción en las inversiones o cambios en las políticas tecnológicas. Sin embargo, a partir de 2020 se percibe una recuperación sostenida, alimentada por la aceleración de la digitalización como respuesta a la pandemia. Este repunte sugiere que el sector del software está jugando un papel cada vez más relevante en la mejora de la productividad y en el fortalecimiento de la economía mexicana.

1 El modelo KLEMS desarrollado por Dale Jorgenson de la Universidad de Harvard en 2010

GRÁFICA 1 México (2002-2022): Tasa porcentual de contribución del software al crecimiento

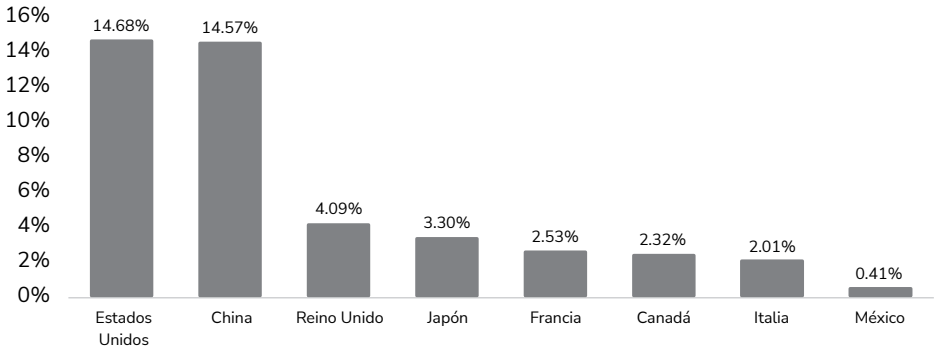


FUENTE: Elaboración propia con datos de INEGI, Productividad de Factores KLEMS

B. IA COMO CREADORA Y DESTRUCTORA DE FICCIONES

La inteligencia artificial, especialmente a través de algoritmos generativos como los modelos de lenguaje, tiene un impacto significativo en la generación narrativa. Estos algoritmos pueden crear relatos futuristas, campañas publicitarias y predicciones económicas que influyen en el imaginario colectivo de varias maneras. Sin embargo, para México, la IA representa tanto una oportunidad como un desafío: el país se encuentra en una encrucijada donde puede convertirse en un líder tecnológico regional o rezagarse en la carrera global. Para que el país pueda aprovechar plenamente estas oportunidades y mantenerse competitivo a nivel global, es crucial aumentar la producción de publicaciones y estudios sobre IA. Una mayor cantidad de investigaciones y publicaciones no solo fomentará la innovación y el desarrollo tecnológico en el país, sino que también posicionará a México como un líder en el campo de la IA. La siguiente gráfica ilustra la necesidad urgente de incrementar las publicaciones sobre IA en México, al compararlo con otros países.

GRÁFICO 2 Países seleccionados (2000 - 2023): Promedio de publicaciones sobre IA
Porcentaje sobre el total de publicaciones



FUENTE: Elaboración propia con datos de AI, Index Report (2023)

CUADRO 1. México publicaciones sobre IA por tema

Tema	Publicaciones
Algoritmos de optimización de inteligencia de enjambre	108
Sistemas y aplicaciones de lógica difusa tipo 2	104
Análisis del impacto económico y del desarrollo urbano	93
Fundamentos y aplicaciones de las redes neuronales	87
Industria 4.0 y transformación digital en la industria manufacturera	85
Inteligencia artificial en medicina	76
Sampling-Based Motion Planning Algorithms	69
Algoritmos de planificación de movimiento basados en muestreo	67
Detección y diagnóstico de fallos de procesos en las industrias	66
Minería de datos educativos y análisis del aprendizaje	61

FUENTE: Elaboración propia con datos de OpenAlex (2024)

Para abordar la necesidad de aumentar las publicaciones sobre inteligencia artificial en México, una solución efectiva sería fomentar la colaboración y el apoyo entre universidades, centros de investigación y el sector privado. Las universidades pueden desempeñar un papel crucial al establecer programas de investigación especializados en IA, ofreciendo incentivos para la publicación de estudios y facilitando el acceso a recursos y financiamiento. Además, promover la creación de redes y consorcios de investigación que incluyan a expertos nacionales e internacionales puede enriquecer la calidad y el alcance de las publicaciones. Implementar estas estrategias no solo incrementará la producción académica en IA, sino que también fortalecerá la posición de México en el panorama global de la tecnología y la innovación.

CUADRO 2 México: Publicaciones sobre inteligencia artificial 2000-2024

Institución	Publicaciones
Instituto Politécnico Nacional	966
Tecnológico de Monterrey	859
Universidad Nacional Autónoma de México	842
Secretaría de Educación Pública	372
Tecnológico Nacional de México	369
Universidad de Guadalajara	359

FUENTE: Elaboración propia con datos de OpenAlex (2024).

Sin embargo, el uso de la inteligencia artificial puede tener consecuencias destructoras significativas, como la erosión de la autenticidad y las predicciones de rendimiento. *La proliferación de deepfakes y narrativas generadas por IA pone en duda la confianza en las fuentes tradicionales de información, ya que se vuelve cada vez más difícil distinguir entre lo real y lo fabricado.* Esto puede socavar la credibilidad de los medios y las instituciones, generando desconfianza generalizada. Además, las decisiones automatizadas basadas en predicciones algorítmicas tienen efectos tangibles en los mercados financieros, las políticas públicas y el comportamiento social. Estas predicciones performativas pueden influir en la toma de decisiones de manera que refuerce ciertos patrones y tendencias, a veces de manera perjudicial, afectando la estabilidad y equidad en diversos ámbitos.

C. LA REESTRUCTURACIÓN ESTRUCTURAL A PARTIR DE LA IA

La inteligencia artificial se presenta como una tecnología con posibilidades infinitas, capaz de revolucionar múltiples industrias al optimizar procesos, crear nuevos modelos de negocio y redefinir el valor de los datos. La IA no solo está transformando el mundo, sino que nos sitúa al borde de una auténtica revolución digital. El avance vertiginoso de la inteligencia artificial ha captado la atención global, generando tanto entusiasmo como preocupación, y desatando un debate crucial sobre su impacto potencial en la economía mundial.

La inversión global en inteligencia artificial ha mostrado un crecimiento significativo en diversas áreas clave, lo que refleja el potencial transformador de esta tecnología en sectores estratégicos. El sector de la atención al cliente y el procesamiento del lenguaje natural ha liderado la inversión, destacando

la importancia de la IA en la mejora de la comunicación y la eficiencia en los servicios. Por su parte, la medicina y la atención sanitaria continúan siendo un área de fuerte inversión, subrayando el impacto de la IA en la optimización de diagnósticos, tratamientos y la gestión de recursos en salud. La gestión y procesamiento de datos también se ha consolidado como un área crucial, evidenciando cómo la capacidad de la IA para analizar grandes volúmenes de datos puede impulsar la toma de decisiones más informadas y precisas.

Además, la infraestructura y la gobernanza de la IA están ganando terreno como áreas esenciales para garantizar el desarrollo ético y sostenible de la tecnología, lo cual es indispensable para abordar los desafíos de transparencia, regulación y seguridad. Otros sectores, como la tecnología educativa y la energía, también se perfilan como áreas de inversión clave, con un enorme potencial para transformar la manera en que se enseña, se aprende y se gestionan los recursos energéticos, impulsando una transición hacia economías más sostenibles y eficientes. Con estos datos en mente, México se encuentra en una posición privilegiada para capitalizar estas tendencias globales, integrando la IA de manera efectiva en sectores como la salud, la educación y la energía, lo cual no solo fomentaría su desarrollo económico, sino que también contribuiría a la creación de un entorno más justo y competitivo a nivel global.

CUADRO 3 Inversión global en Inteligencia Artificial 2013-2023

Miles de millones de dólares a precios constantes 2021

Total	**568.65**
Procesamiento del lenguaje natural y atención al cliente	48.05
Medicina y atención sanitaria	33.21
Gestión y procesamiento de datos	22.27
Infraestructura y gobernanza de la IA	17.82
Tecnología educativa	12.47
Energía, petróleo y gas	12.25
Vehículos autónomos	11.69
Tecnología financiera	9.68
Contenido creativo, musical y de vídeo.	9.25
Semiconductores	9.18
Reconocimiento facial	9.04
Drones	7.62
Tecnología de seguros	5.09
Capital de riesgo	4.71
Ciberseguridad y protección de datos	4.62

FUENTE: Elaboración propia con datos de Quid (2024).

México ha consolidado su liderazgo en el crecimiento de empresas de inteligencia artificial en América Latina, con un aumento del 965% en los últimos seis años, alcanzando 362 empresas en 2024. Este crecimiento ha generado más de 11,000 empleos y atraído inversiones superiores a los 500 millones de dólares. Además, la Ciudad de México se ha convertido en el principal centro de desarrollo de IA, albergando casi la mitad de las startups del país. Las empresas en este sector, que incluyen áreas clave como el aprendizaje automático y la robótica, reportan ingresos promedio de 1.3 millones de dólares en 2024, lo que refleja un mercado en expansión y con gran potencial.

CUADRO 4 Empresas de IA en América Latina 2024

País	2018	2024	Tasa de crecimiento
Brasil	124	728	487%
México	34	362	965%
Chile	28	160	471%
Colombia	13	100	669%
Argentina	32	83	159%

FUENTE: https://www.santander.com.mx/ceb/2024/mexico-lidera-crecimiento-de-empresas-de-inteligencia-artificial-en-latam.html

D. SECTORES CON ALTO POTENCIAL PARA LA IA

I. Salud

La inteligencia artificial ya no es solo una promesa futurista, sino que ya es una realidad. La IA está transformando el sector salud, abriendo nuevas oportunidades y desafíos. Un ejemplo que ilustra este impacto es el trabajo de Demis Hassabis y John Jumper de Google DeepMind, junto con David Baker, quienes ganaron el Premio Nobel de Química 2024. Su creación, AlphaFold, logró predecir las estructuras tridimensionales de proteínas a partir de secuencias de aminoácidos, resolviendo un desafío científico de más de 50 años. Este avance permitió identificar la estructura de casi 200 millones de proteínas, proporcionando herramientas para abordar problemas como la resistencia a los antibióticos y el diseño de enzimas capaces de descomponer plásticos. Además, Baker avanzó aún más, desarrollando proteínas completamente nuevas con aplicaciones que abarcan desde la salud hasta la sostenibilidad (Expansión Digital, 2023).

La IA ya está demostrando su capacidad para transformar diagnósticos y tratamientos. Por ejemplo, algoritmos avanzados entrenados con más de 60,000 mamografías ahora pueden identificar patrones sutiles en tejidos mamarios, detectando posibles tumores malignos en etapas extremadamente tempranas. Este desarrollo promete erradicar los diagnósticos tardíos y revolucionar la detección temprana del cáncer (National Geographic, 2023).

Además, la IA tiene un papel clave en regiones remotas donde el acceso a médicos es limitado. Herramientas de diagnóstico basadas en IA pueden ofrecer consultas precisas y recomendaciones de tratamiento sin la necesidad de presencia médica inmediata. Sin embargo, este avance plantea preguntas sobre el futuro de ciertos roles médicos, como el de los médicos generales, quienes podrían ser parcialmente reemplazados. Hay una gran oportunidad para mejorar la detección de enfermedades, como señalan Kreutzer y Sirrenberg, el *Camelyon Grand Challenge 2016* demostró que la colaboración entre humanos y algoritmos puede generar mejores resultados. En este desafío, un algoritmo desarrollado por investigadores de Harvard y MIT logró una precisión del 92.5% en la detección de cáncer de mama metastásico, mientras que los patólogos alcanzaron un 96.6%. La combinación de ambos enfoques resultó en una precisión del 99.5%, reduciendo el margen de error en un 85%.

En el futuro, los dispositivos portátiles equipados con IA podrían cambiar radicalmente la forma en que las personas gestionan su salud. Estos dispositivos serían capaces de recopilar datos en tiempo real, recomendar dietas, sugerir medicamentos personalizados y ofrecer consejos adaptados a las necesidades específicas de cada individuo, mejorando el bienestar general.

La implementación de IA no solo se limitará al diagnóstico y tratamiento. También promete optimizar la gestión hospitalaria, administrando recursos de manera más eficiente, mejorando el flujo de pacientes y reduciendo los tiempos de espera. Asimismo, en la distribución de medicamentos, la IA podría asignar recursos de manera más precisa, personalizando tratamientos para maximizar resultados.

II. Finanzas

El sector financiero es una de las industrias con mayor potencial para beneficiarse de la implementación de tecnologías basadas en inteligencia artificial. Históricamente, este sector ha liderado la adopción tecnológica, siendo pionero en el uso de computadoras para innovaciones como el comercio de alta frecuencia.

Según IBM (2023), la inteligencia artificial en las finanzas se fundamenta en la aplicación de tecnologías avanzadas, especialmente algoritmos de aprendizaje automático. Estas herramientas permiten a las instituciones financieras optimizar la eficiencia, la precisión y la velocidad en tareas como el análisis de datos, las proyecciones, la gestión de inversiones y riesgos, la detección de fraudes y la atención al cliente. La IA está transformando el sector al automatizar procesos bancarios tradicionales y generar interacciones con los clientes que imitan la inteligencia y el comportamiento humano.

Plummer (2024) señala que la IA también está revolucionando la detección y prevención del fraude. Los modelos de aprendizaje automático pueden analizar millones de transacciones para identificar patrones que indiquen actividades fraudulentas. Los bancos aprovechan estas capacidades para identificar transacciones sospechosas en tiempo real, reduciendo significativamente las pérdidas por fraude, así como la personalización de servicios. Los bancos utilizan análisis de datos avanzados para ofrecer recomendaciones de productos adaptadas a los hábitos de gasto e historial de transacciones de los clientes.

Además, Plummer también menciona como la IA permite realizar análisis financieros y modelos de riesgo más sofisticados. Al identificar correlaciones en vastos conjuntos de datos que no serían evidentes para los humanos, la IA mejora los análisis predictivos, la planificación de escenarios y las evaluaciones de riesgo.

La IA tiene el potencial de ofrecer muchos beneficios al sector financiero, como la optimización en la gestión de carteras, el análisis eficiente de grandes volúmenes de datos, predicciones más precisas de rentabilidad, asignación óptima de calificaciones crediticias y evaluaciones de riesgos más exactas. Posicionando a la IA como motor del futuro del sector financiero.

III. Manufactura

La inteligencia artificial ha impulsado la llamada Industria 4.0, donde los procesos productivos están cada vez más automatizados y conectados, el cual tiene un impacto en la mejora operativa y la reducción de costos.

La IA ha revolucionado la industria al impulsar el mantenimiento productivo, la robótica avanzada y la optimización de la cadena de suministro. Existen sensores que monitorean equipos en tiempo real para prevenir fallas, mientras que robots con IA trabajan junto a humanos en líneas de ensamblaje, incrementando la productividad. Además, herramientas como las utilizadas por Amazon predicen la demanda, ajustan inventarios y optimizan las rutas de entrega. Estos avances han elevado la eficiencia productiva, reduciendo costos y permitiendo la mejora en los procesos de producción.

La industria automotriz es uno de los principales sectores que ha adoptado la inteligencia artificial y la robótica en México, además es de las industrias más importantes del país. Las empresas automotrices la utilizan para optimizar las cadenas de suministro, mejorar la planificación de la producción y detectar fallas en vehículos antes de que ocurran. La inteligencia artificial permite el análisis predictivo, el cual reduce el tiempo ineficiente y mejora la calidad de los productos. Además, los robots colaborativos trabajan junto a los empleados para aumentar la eficiencia en el proceso de producción.

El sector electrónico también ha integrado soluciones de IA para optimizar sus procesos productivos. Los sistemas de aprendizaje automático identifican cuellos de botella en la producción y ajustan los flujos de trabajo, maximizando la productividad. La visión por computadora se utiliza para el control de calidad, permitiendo una detección más precisa y rápida de defectos en productos electrónicos.

En el sector alimentario, las empresas están implementando IA para mejorar la eficiencia operativa y garantizar estándares de calidad. Los mecanismos de automatización y los algoritmos de aprendizaje automático ayudan a gestionar inventarios y prever demandas, optimizando la producción y reduciendo costos. Además, las tecnologías de IA permiten una inspección automatizada durante el proceso de fabricación, mejorando la calidad de la producción final.

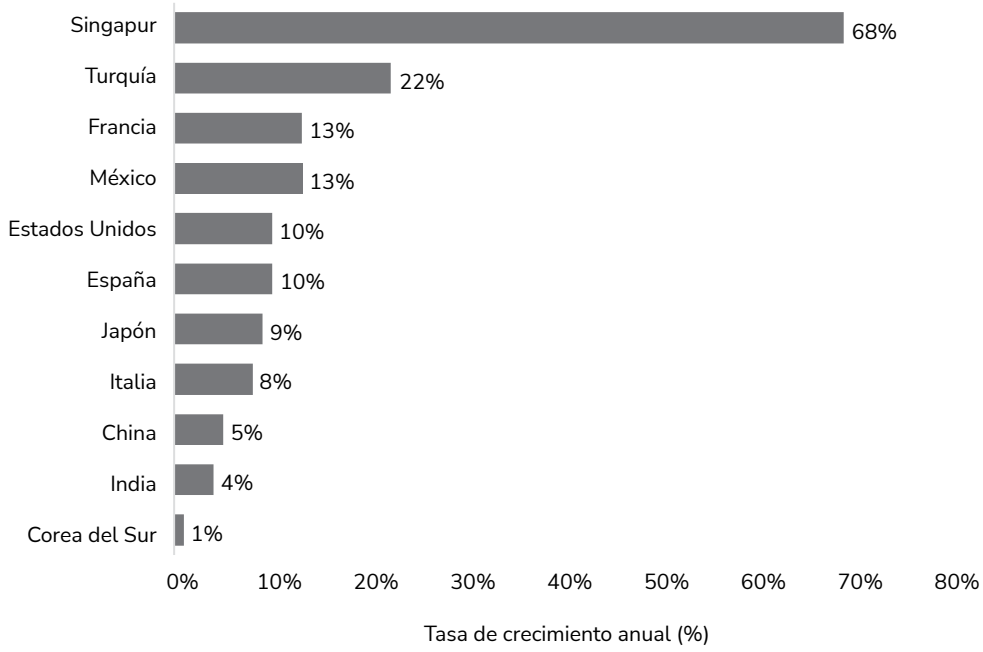
La industria farmacéutica también ha adoptado las tecnologías de la IA para mejorar sus procesos productivos. La utilizan más que nada para el análisis de datos relacionados con el desarrollo de nuevos medicamentos, optimizando el proceso de investigación y desarrollo.

El sector textil ha implementado soluciones basadas en IA para personalizar productos a gran escala. Esto no solo mejora la satisfacción de los consumidores al adaptar diseños a tendencias específicas, sino que también optimiza el uso de los recursos, reduciendo costos y aumentando la eficiencia.

La industria manufacturera se está beneficiando enormemente de la inteligencia artificial generativa, especialmente en áreas como el diseño de productos, la optimización de procesos y la automatización. Según un informe de Stack Overflow[2] (2023), el 76% de los profesionales en la industria del software ya utiliza o planea utilizar herramientas de IA en su trabajo. Esta tendencia también se refleja en el sector manufacturero, donde un número creciente de empresas adopta tecnologías de IA para mejorar la productividad y reducir los costos operativos.

2 Stack Overflow 2023 Developer Survey

GRÁFICO 3. Tasa de crecimiento anual de robots industriales instalados por país, 2021 vs. 2022



NOTA: Elaboración propia con datos de: IFR 2023, AI Index Report

El 82% de los profesionales encuestados destacan que los principales beneficios de utilizar herramientas de IA incluyen un aumento significativo en la productividad, al permitir una mayor velocidad en el diseño y la producción. Alrededor del 60% de los encuestados también reporta una mayor eficiencia y un aprendizaje acelerado, especialmente en la aplicación de IA para la automatización de tareas repetitivas y en la mejora de la calidad del producto. Estas herramientas son especialmente útiles en la simulación de procesos de manufactura y la optimización de la cadena de suministro.

A pesar de estos progresos, persisten desafíos críticos, como la validez de los datos en aplicaciones reales, la transparencia e interpretabilidad de los modelos de IA y la necesidad de integrar enfoques físicos con métodos basados en datos. También se identifican oportunidades para el desarrollo de IA generativa con control granular y sistemas generalistas que interactúen de manera más natural con los usuarios. El futuro de la manufactura dependerá de la adopción de nuevas herramientas y metodologías que permitan aprovechar ambientes ricos en datos y comprender a fondo los procesos industriales. Este enfoque

no solo mejorará la eficiencia operativa, sino que también consolidará las bases científicas para avanzar hacia una manufactura verdaderamente inteligente (Gao *et al*.. 2024).

IV. Agricultura

Los desperdicios de alimentos es un problema persistente que afecta tanto a la economía como al medio ambiente. No se trata solo de lo que queda en el plato, sino de una pérdida masiva que ocurre a lo largo de toda la cadena de suministro, desde la producción hasta la distribución. En Estados Unidos, el USDA estima que entre el 30% y el 40% del suministro de alimentos se desperdicia, un porcentaje alarmante que refleja ineficiencias en las diferentes etapas del proceso. Estas pérdidas pueden deberse al deterioro durante el transporte, condiciones inadecuadas de almacenamiento o incluso la sobreproducción.

Sin embargo, la ciencia de datos y la inteligencia artificial están comenzando a cambiar esta realidad. En la producción agrícola, tecnologías como drones, sensores y satélites recopilan datos sobre la productividad del suelo, la salud de los cultivos y las condiciones climáticas. Procesados en tiempo real, estos datos permiten a los agricultores tomar decisiones más inteligentes sobre rotación de cultivos, siembra y cosecha. Además, algoritmos predictivos pueden anticipar plagas o enfermedades, facilitando intervenciones tempranas y eficaces. Robots agrícolas autónomos complementan estas tareas al sembrar, cosechar y aplicar pesticidas con precisión, reduciendo desperdicios y optimizando recursos.

En la etapa de distribución, la IA también juega un papel clave. Sensores conectados a sistemas analíticos avanzados monitorean parámetros críticos como la temperatura, activando alertas en caso de variaciones que podrían comprometer la calidad de los productos. En el comercio minorista, sistemas automatizados de seguimiento de fechas de caducidad ayudan a planificar promociones y evitar que los alimentos pierdan frescura. Por otro lado, algoritmos de predicción ajustan los pedidos de productos según la demanda, alineando oferta y consumo para minimizar el desperdicio. Estas tecnologías no solo mejoran la eficiencia, sino que también contribuyen a construir una cadena de suministro más sostenible y consciente. (Innovación Digital 360, 2024)

V. Comercio minorista

La IA ha transformado la experiencia de compra y las operaciones logísticas. Además, los sistemas predictivos de gestión de inventarios optimizan los niveles de existencias, reduciendo el desperdicio y asegurando que los productos estén disponibles cuando se necesitan. Un ejemplo innovador de la aplicación

de IA es Amazon Go, donde la tecnología permite una experiencia de compra completamente automatizada sin necesidad de cajeros.

Los algoritmos de IA analizan el comportamiento de compra y sugieren productos personalizados, como lo hace Amazon del mismo modo que el ejemplo anterior, Amazon. Esto aumenta la probabilidad de ventas cruzadas y repetidas.

Algo que ya la mayoría de los minoristas y empresas grandes han comenzado a implementar son los chatbots basados en IA, estos pueden responder preguntas frecuentes de los clientes y mejorar la atención al cliente las 24 horas del día.

Estas innovaciones han permitido crear experiencias de compra más personalizadas, mejorar la eficiencia y desarrollar nuevos modelos de comercio sin fricciones, beneficiando tanto a los consumidores como a los minoristas.

VI. Educación

La inteligencia artificial está revolucionando la forma en que aprendemos y enseñamos. En el último año, muchas plataformas académicas como Khan Academy y Coursera, reconocidas por ofrecer videos, cursos y lecturas, han comenzado a incorporar chatbots dentro de sus páginas diseñados para facilitar el aprendizaje. Estas herramientas pueden generar cuestionarios personalizados, resumir contenidos y proporcionar temarios adaptados a las necesidades de cada usuario.

El impacto potencial de la IA en la educación va mucho más allá. Con la capacidad de crear imágenes y videos, el futuro podría ofrecernos clases completas impartidas por profesores virtuales generados por IA, capaces de escribir en pizarras como si fueran docentes reales. Estas clases serían completamente personalizadas, ajustándose al ritmo de cada estudiante.

Los sistemas de inteligencia artificial pueden identificar las fortalezas y debilidades de cada estudiante y adaptar el contenido a sus necesidades, permitiendo que cada uno aprenda a su propio ritmo. Esto no se queda hasta aquí, sino que también hay una forma de apoyar a los docentes y mejorar de esta forma su rendimiento, por ejemplo, la IA puede encargarse de calificar exámenes de opción múltiple, organizar datos de los estudiantes o programar actividades, liberando tiempo para que los docentes se enfoquen en enseñar.

CUADRO 5. México: publicaciones sobre IA por tema

Tema	Publicaciones
Algoritmos de optimización de inteligencia de enjambre	108
Sistemas y aplicaciones de lógica difusa tipo 2	104
Análisis del impacto económico y del desarrollo urbano	93
Fundamentos y aplicaciones de las redes neuronales	87
Industria 4.0 y transformación digital en la industria manufacturera	85
Inteligencia artificial en medicina	76
Sampling-Based Motion Planning Algorithms	69
Algoritmos de planificación de movimiento basados en muestreo	67
Detección y diagnóstico de fallos de procesos en las industrias	66
Minería de datos educativos y análisis del aprendizaje	61

FUENTE: Elaboración propia con datos de OpenAlex (2024)

Más allá de facilitar el acceso al conocimiento, la IA está abriendo las puertas a una experiencia educativa más inclusiva, personalizada y eficiente. Esto tiene el potencial de reducir las brechas educativas en regiones del mundo con acceso limitado a una educación de calidad. La implementación de estas tecnologías podría ser clave para cerrar las desigualdades, impulsar el desarrollo en países con rezago educativo y generar nuevas oportunidades para millones de personas.

VII. Entretenimiento y medios

La inteligencia artificial ha transformado tanto la creación de contenido como la forma en que los usuarios lo consumen, por ejemplo, ChatGPT permite la creación automatizada de guiones, imágenes y contenido audiovisual, facilitando la producción creativa a una escala sin precedentes. Además, plataformas como Spotify y YouTube utilizan IA para personalizar las recomendaciones de música, videos y podcasts, ofreciendo a los usuarios contenido adaptado a sus gustos y preferencias. En la industria del cine, la IA optimiza la producción al predecir el éxito de las películas y gestionar los presupuestos de manera más eficiente. Estas innovaciones no solo han democratizado la creación de contenido, sino que también han mejorado la experiencia del usuario al hacerla más personalizada y accesible.

Esto ha afectado y beneficiado a las grandes empresas, por ejemplo a Walt Disney Company ya que está transformando diversas áreas del entretenimiento:

en la producción de contenidos, como en *The Mandalorian*, mejora los efectos visuales y la narrativa al rejuvenecer personajes; en parques temáticos, como *Pirates of the Caribbean*, facilita animatrónicos interactivos que ofrecen experiencias personalizadas; en películas animadas, como *Elemental*, aporta realismo y eficiencia al simular elementos como fuego y agua. Además, en estrategias de contenido, optimiza procesos y reduce costos mediante herramientas de posproducción, y en experiencias inmersivas, enriquece la interacción en parques temáticos con aplicaciones de realidad aumentada.

El uso de la IA de la industria cinematográfica y el streaming también ha tenido consecuencias negativas, a causa de la reciente huelga sindical SAG-AFTRA, la cual comenzó el 14 de julio de 2023 y concluyó el 8 de noviembre tras 118 días de paro. Fue una de las más largas en la historia de la industria. Durante este período, la economía californiana sufrió pérdidas estimadas en 6,500 millones de dólares, además del despido de unos 45,000 trabajadores. Uno de los motivos fue la creciente preocupación por el uso de IA para replicar voces o apariencias sin compensación adecuada a los actores.

VIII. Energía

La inteligencia artificial y la industria 4.0 exigen un cambio en el patrón energético de la humanidad y los procesos productivos. Con la adopción de la inteligencia artificial, las empresas tecnológicas han visto un aumento de las emisiones debido a la creciente demanda de energía de la IA, en particular de los centros de datos necesarios para entrenar y ejecutar modelos de IA. A medida que *los sistemas de IA generativa ganan popularidad, su consumo de energía aumenta y algunos modelos requieren mucha más energía que el software tradicional. A medida que los sistemas de IA se hacen más avanzados, se espera que su consumo de energía se duplique aproximadamente cada 100 días, ejerciendo presión sobre unas redes eléctricas ya de por sí sobrecargadas.* Empresas como Microsoft y Google han informado de aumentos significativos en las emisiones de CO^2, en parte debido a la expansión de sus centros de datos (Kemene *et al*, 2024).

A pesar de la elevada demanda energética de la IA, ésta tiene potencial para impulsar la transición energética optimizando la integración de las energías renovables, mejorando la eficiencia energética en las industrias intensivas en carbono y contribuyendo a la estabilidad de la red. La IA puede ayudar a predecir la producción de energía, cambiar las cargas energéticas y optimizar el uso de fuentes de energía renovables, garantizando un funcionamiento eficiente de la red. Sin embargo, la energía necesaria para apoyar el crecimiento de la IA plantea el reto de equilibrar sus beneficios con los costes

medioambientales. A medida que el Parlamento Europeo y otros organismos reguladores empiezan a establecer requisitos para que los sistemas de IA sean eficientes desde el punto de vista energético, los avances en hardware, como el nuevo «superchip» de Nvidia, ofrecen esperanzas para reducir el consumo de energía (Kemene *et al.*, 2024).

La IA ha tenido un impacto significativo en el sector energético, promoviendo la sostenibilidad y la eficiencia. Los sistemas inteligentes de gestión de redes eléctricas equilibran la oferta y la demanda en tiempo real, asegurando un suministro estable y eficiente. Los algoritmos de predicción climática analizan datos meteorológicos para optimizar la generación de energía renovable, maximizando su uso y minimizando el desperdicio. Además, los edificios inteligentes utilizan IA para ajustar automáticamente la iluminación y la temperatura, mejorando la eficiencia energética. Estas innovaciones han contribuido a la reducción de emisiones de carbono y a una mayor integración de energías limpias, avanzando hacia un futuro más sostenible.

Los sistemas de energía, antes centralizados, ahora requieren flujos de electricidad multidireccionales entre generadores distribuidos, la red y los usuarios, lo que incrementa la complejidad de la gestión. Este fenómeno se ve acentuado por la creciente interconexión entre la energía, el transporte, la construcción y la industria. Por esta razón, la IA se posiciona como una herramienta clave para mejorar la predicción de la oferta y la demanda, optimizar la integración de energías renovables y aumentar la eficiencia operativa. Por ejemplo, el uso de modelos de aprendizaje automático para prever la producción eólica ha permitido a empresas como Google aumentar la precisión de sus pronósticos, mejorando así la rentabilidad de sus parques eólicos y permitiendo una venta anticipada de energía a precios más altos (EIA, 2024).

Además, la IA también juega un papel crucial en la prevención de fallas en las redes mediante el mantenimiento predictivo, lo que reduce significativamente las interrupciones en el servicio. Empresas como E.ON y Enel han implementado algoritmos de aprendizaje automático para detectar problemas en sus infraestructuras, anticipándose a fallas antes de que ocurran. A medida que la digitalización y la automatización se expanden, la IA promete revolucionar la gestión de las redes, facilitando la respuesta a la demanda, mejorando el control del flujo de electricidad y ofreciendo servicios al consumidor más eficientes (IEA, 2024).

IX. Transporte y logística

El transporte autónomo y la optimización de rutas están liderados por la IA. La inteligencia artificial ha revolucionado el sector del transporte, mejorando la seguridad, reduciendo costos operativos y aumentando la eficiencia en el transporte urbano. Empresas como Tesla y Waymo están a la vanguardia en el desarrollo de vehículos autónomos, que permiten la conducción sin intervención humana, prometiendo un futuro con menos accidentes y mayor comodidad. En el ámbito de la logística, compañías como UPS utilizan IA para optimizar rutas, lo que reduce el tiempo y los costos de entrega, mejorando la eficiencia operativa. Además, las ciudades inteligentes implementan sistemas de gestión de tráfico basados en IA para regular los flujos vehiculares en tiempo real, aliviando la congestión y mejorando la movilidad urbana. Estas innovaciones están transformando el transporte, haciéndolo más seguro, económico y eficiente.

La Agencia Internacional de la Energía (2024) estima que casi uno de cada cinco vehículos vendidos en 2023 fue eléctrico, lo que representa el doble de la proporción registrada hace apenas dos años. En cuanto al impacto de la inteligencia artificial en el transporte, el avance hacia los vehículos eléctricos (VE) es un componente clave de la transición hacia un sistema de transporte más sostenible, con la IA desempeñando un papel crucial en la optimización de la infraestructura de carga, la gestión de la flota de vehículos eléctricos y la mejora de la eficiencia operativa en este sector.

En México, la fabricación de equipo de transporte es la actividad económica con el mayor valor agregado, y la integración de IA en la industria automotriz podría transformar aún más este sector clave para la economía del país. A medida que las tecnologías emergentes, como la IA, se incorporan en los procesos de producción, México tiene la oportunidad de consolidarse como un líder en la fabricación de vehículos de próxima generación, incluidos los vehículos eléctricos (VE) y autónomos.

D. ALGUNAS REFLEXIONES HACIA EL FUTURO

Aunque la inteligencia artificial tiene un impacto transformador, plantea desafíos éticos y sociales, como la posible pérdida de empleos, la desigualdad en el acceso a estas tecnologías y la dependencia de sistemas centralizados. Podemos abordar ambos temas de manera complementaria. Primero, exploraremos corrientes alternativas al capitalismo que buscan incorporar o resistir la tecnología, incluida la IA, y luego examinaremos casos específicos de transformación industrial con IA.

Corrientes alternativas al capitalismo y su relación con la IA. Algunas corrientes contemporáneas están replanteando los modelos económicos dominantes y, en particular, la relación con tecnologías disruptivas como la IA. Estas perspectivas buscan aprovechar estas herramientas de manera inclusiva, ética y sostenible:

A. **Economía del bien común.** Propuesta por Christian Felber[3], esta corriente promueve un modelo económico basado en valores como cooperación, sostenibilidad y equidad. En este contexto, la IA podría ser usada para: optimizar la distribución de recursos esenciales, como agua, alimentos y energía; analizar datos sociales y ambientales para tomar decisiones políticas más responsables; crear plataformas transparentes de gobernanza colectiva. **Ejemplo de uso de IA**: Algoritmos para la optimización de cultivos en cooperativas agrícolas que trabajan bajo principios de sostenibilidad.

B. **Decrecimiento.** El decrecimiento cuestiona la obsesión con el crecimiento económico infinito, proponiendo reducir el consumo para preservar los recursos del planeta. La IA puede jugar un papel en: diseñar sistemas energéticos más eficientes que prioricen energías renovables; modelar y prever los impactos ecológicos de decisiones políticas o industriales; impulsar economías locales mediante plataformas descentralizadas, reduciendo intermediarios corporativos. **Ejemplo de uso de IA:** Simulaciones de consumo de recursos naturales para comunidades que buscan minimizar su huella ecológica.

C. **Economía circular.** Este modelo se basa en reutilizar materiales y productos para minimizar el desperdicio. La IA ayuda en: diseñar productos más duraderos y reciclables; crear sistemas de logística inversa que recuperen materiales usados; monitorizar procesos industriales para evitar desperdicio de recursos. **Ejemplo de uso de IA**: Plataformas como Winnow utilizan IA para reducir desperdicios de alimentos en restaurantes, maximizando la reutilización.

D. **Propuestas postcapitalistas.** Movimientos como el **"solarpunk"** o el **"procomún digital"** imaginan futuros que combinan tecnología avanzada (como la IA) con estructuras sociales cooperativas:
   - **Solarpunk:** Integra tecnologías como la IA en comunidades autosuficientes, con énfasis en sostenibilidad y diseño ecológico.
   - **Procomún digital:** Recursos abiertos y gestionados colectivamente, como software de IA para investigación o educación.

3 Felber, Christian. La economía del bien común. Barcelona: Ediciones Deusto, 2012.

Ejemplo de uso de IA: Proyectos de código abierto como Hugging Face democratizan el acceso a herramientas de IA, alejándose del control corporativo.

E. CASOS ESPECÍFICOS DE TRANSFORMACIÓN INDUSTRIAL CON IA

La IA también está remodelando industrias clave bajo lógicas capitalistas, pero con potencial para cambiar la forma de operar.

La inteligencia artificial está remodelando industrias clave bajo lógicas capitalistas, con un potencial significativo para transformar sus formas tradicionales de operación. En el ámbito del transporte autónomo y logística, empresas como Tesla y Waymo lideran el desarrollo de vehículos sin conductor, revolucionando el sector al reducir costos operativos y eliminar la necesidad de conductores humanos, lo que además genera un impacto ambiental positivo mediante la optimización de rutas y la consecuente reducción de emisiones de carbono.

En agricultura sostenible, compañías como Blue River Technology, filial de John Deere, utilizan inteligencia artificial para identificar y aplicar herbicidas de manera selectiva, lo que disminuye el uso de químicos y costos. En África, proyectos como Hello Tractor emplean IA para gestionar tractores compartidos, permitiendo que pequeños agricultores mejoren significativamente su productividad.

La creación de contenido cultural experimenta una democratización gracias a plataformas como Runway ML, que permiten a artistas generar imágenes y videos mediante IA, y Deepdub, que usa tecnología inteligente para traducir y doblar películas a múltiples idiomas, ampliando la difusión cultural.

En el ámbito médico y de accesibilidad global, organizaciones como Butterfly Network desarrollan dispositivos de ultrasonido portátiles impulsados por IA para realizar diagnósticos en áreas remotas. En India, herramientas inteligentes traducen contenido médico a idiomas locales, facilitando el acceso a comunidades rurales.

Finalmente en el sector energético, empresas como Google DeepMind[4] han logrado la predicción de generación eólica, optimizando el uso de energías renovables en redes eléctricas.

4 Elkin, Carl, y Sims Witherspoon. "Machine Learning Can Boost the Value of Wind Energy." DeepMind, 26 de febrero de 2019. Consultado el 16 de Septiembre de 2024. https://deepmind.google/discover/blog/machine-learning-can-boost-the-value-of-wind-energy/

V. CONCLUSIÓN Y REFLEXIONES FINALES

La inteligencia artificial está en un momento decisivo que podría definir el futuro de nuestras sociedades. Hoy en día, su desarrollo está principalmente impulsado por intereses económicos que buscan maximizar ganancias y consolidar monopolios tecnológicos. Sin embargo, no todo está escrito. La IA también tiene el poder de convertirse en una herramienta para construir un mundo más justo e inclusivo. Eso sí, este cambio no ocurrirá por arte de magia. Necesitamos un marco ético, político y social que oriente su implementación hacia objetivos que beneficien al bienestar colectivo, no solo a los intereses de unos pocos.

Hacen falta reglas éticas sólidas que limiten los abusos de las grandes corporaciones y garanticen un uso responsable de la IA. Estas normas deben abordar cuestiones como la privacidad, la transparencia y la equidad, previniendo que el poder tecnológico agrave las desigualdades existentes. Segundo, democratizar la IA es imprescindible. No puede ser una herramienta exclusiva para quienes tienen los recursos; debe estar al alcance de todos, sin importar su situación económica o dónde vivan. Esto significa reducir las brechas digitales y fomentar la educación tecnológica en comunidades que suelen quedar marginadas.

Por último, es hora de repensar el papel de la tecnología en nuestras vidas. En lugar de perpetuar un modelo centrado únicamente en el crecimiento económico, necesitamos construir un futuro basado en la sostenibilidad, la equidad social y el respeto por el medio ambiente. La IA puede ser un motor para lograr todo esto, pero solo si tomamos decisiones conscientes, con una visión a largo plazo. Así que combinar la innovación tecnológica con valores humanistas no es solo importante; es esencial. De esta manera, podemos asegurarnos de que la tecnología esté verdaderamente al servicio de la humanidad, y no al revés.

VI. BIBLIOGRAFÍA

Argandoña, A. "Desafíos y oportunidades de la IA en el sistema financiero mexicano." En Transformación digital en América Latina, 73–98. Barcelona: IESE Business School, 2019.

Beckert, Jens. Imagined Futures: Functional Expectations and Capitalist Dynamics. Cambridge, Massachusetts: Harvard University Press, 2016.

Crespo Garay, Cristina. "Un método basado en inteligencia artificial detecta el cáncer del futuro." National Geographic Latinoamérica, 10 de abril de 2023. https://www.nationalgeographicla.com/ciencia/2023/04/un-metodo-basado-en-inteligencia-artificial-detecta-el-cancer-del-futuro.

Endeavor México. La Era de la IA en México. 2024. https://mexico.endeavor.org/2024-ia/.

Expansión Digital. "Científicos de Google usan la IA y ganan el Premio Nobel de Química 2024." Expansión, 10 de octubre de 2024. https://expansion.mx/mundo/2024/10/09/premio-nobel-de-quimica-2024.

Forbes Staff. "Inteligencia artificial, ¿la vía para transformar la economía mexicana?" Forbes México, 2024.

Franco, G. "Disney pone en marcha una iniciativa con Inteligencia Artificial que transformaría Hollywood para siempre." Hipertextual, 25 de octubre de 2024. https://hipertextual.com/2024/10/disney-pone-en-marcha-iniciativa-inteligencia-artificial.

Gao, Robert X., Qing Chang, Jorge F. Arinez, Chengying Xu, y Jianjing Zhang. "Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook." Journal of Manufacturing Science and Engineering 142, no. 11 (2024): 110804.

Gao, Robert X., Jörg Krüger, Marion Merklein, Hans-Christian Möhring, y József Váncza. "Artificial Intelligence in Manufacturing: State of the Art, Perspectives, and Future Directions." CIRP Annals 73, no. 2 (2024): 723–49. https://doi.org/10.1016/j.cirp.2024.04.101.

Gómez Mendoza, C., y Ezar, R. "Manufactura avanzada en la era de la inteligencia artificial: ¿Qué relevancia tienen los datos para esta industria?" EY México, 2024.

IBM. "¿Qué es la inteligencia artificial en las finanzas?" Por Matthew Finio y Amanda Downie. Última modificación: 8 de diciembre de 2023. Consultado el 19 de noviembre de 2024. https://www.ibm.com/es-es/topics/artificial-intelligence-finance.

IEA. Global EV Outlook 2024. París: IEA, 2024. https://www.iea.org/reports/global-ev-outlook-2024.

IEA. "Why AI and energy are the new power couple." Última modificación: 2023. Consultado el 19 de noviembre de 2024. https://www.iea.org/commentaries/why-ai-and-energy-are-the-new-power-couple.

Innguma. "Disney y la IA: Reinventando la innovación y vigilancia tecnológica." Última modificación: 21 de febrero de 2024. https://www.innguma.com/disney-ia-innovacion-vigilancia-tecnologica/.

Innovación Digital 360. "Desperdicio de alimentos: cómo reducirlo con IA y ciencia de datos." Última modificación: 14 de julio de 2022. https://www.innovaciondigital360.com/agrotech/desperdicio-de-alimentos-como-reducirlo-con-ia-y-ciencia-de-datos/.

Kemene, Eleni, Bart Valkhorf, y Thapelo Tladi. "AI and energy: Will AI help reduce emissions or increase demand? Here's what to know." World Economic Forum, 2024.

Keywords/-/Meta/revistaeyn-Com. "Disney forma una unidad para coordinar el uso de IA y realidad aumentada." Revista E&N, 2 de noviembre de 2024. https://www.revistaeyn.com/empresasymanagement/disney-forma-una-unidad-para-coordinar-el-uso-de-ia-y-realidad-aumentada-FE22405167.

Kreutzer, Ralf, y Marie Sirrenberg. Understanding Artificial Intelligence: Fundamentals, Use Cases and Methods for a Corporate AI Journey. Cham: Springer, 2019.

Palma, J., y Alvarado, R. "Impacto de la inteligencia artificial en México: Perspectivas éticas y sociales." Revista Misión Jurídica, 2015.

Plummer, S. "La IA en las finanzas: Revolucionar el futuro de la gestión financiera." El Experto en Anualidades, 11 de septiembre de 2024. Consultado el 19 de noviembre de 2024. https://www.datacamp.com/es/blog/ai-in-finance.

UNESCO. Propuesta de Agenda Nacional de la Inteligencia Artificial para México 2024-2030. 2024.

USDA. "Food Waste FAQs." Consultado el 1 de noviembre de 2024. https://www.usda.gov/foodwaste/faqs.

INTELIGENCIA ARTIFICIAL Y LAS LÍNEAS DE CONSTRUCCIÓN DE UNA NUEVA ECONOMÍA DIGITAL

SUMARIO

I. Introducción: el cambio de modelo productivo ante la Inteligencia Artificial **II.** Efectos de la IA en el futuro laboral **III.** Incremento de concentración a partir de la digitalización y de la IA **IV.** Acumulación de capacidades de desarrollo tecnológico y afectación al proceso de desarrollo **V.** Impulso a la Inteligencia Artificial y sustentabilidad **VI.** Elementos para una agenda económica en materia de IA **VII.** A modo de conclusión **VIII.** Bibliografía

> JULEN BERASALUCE IZA

Profesor-Investigador del Centro de Estudios
Económicos de El Colegio de México (COLMEX)
Consejero del Instituto para el Desarrollo Industrial
y la Transformación Digital (INADI)

"Insistes en que hay cosas que las máquinas no pueden hacer.
Si tú me dices exactamente qué es lo que no pueden hacer,
yo siempre seré capaz de construir una máquina que haga exactamente eso"
JOHN VON NEUMANN, MANIAC

I. INTRODUCCIÓN: EL CAMBIO DE MODELO PRODUCTIVO ANTE LA INTELIGENCIA ARTIFICIAL

Durante las últimas décadas, estamos viviendo una revolución tecnológica marcada por el auge de las tecnologías digitales, que tiene su expresión en todas las dimensiones del orden socioeconómico: interacción social, transmisión de la información, ocio, productividad, etc. Los diferentes conceptos empleados, como Industria 4.0 (Alemania)o Industria 5.0 (Comisión Europea), Sociedad 5.0 (Japón), refieren bajo diferentes acercamientos a nuevas concepciones sociales que integran protagónicamente las tecnologías digitales. En ejercicios que combinan el análisis de las más recientes tecnologías digitales y de prospectiva sobre su desarrollo, muchos países (desarrollados y en desarrollo) están tratando de adaptarse a los cambios venideros, en el entendido de que la profundidad del cambio socioeconómico vinculado a la Revolución Digital exige una mirada prospectiva hacia los beneficios y riesgos de las nuevas tecnologías.

Dentro de las tecnologías digitales más disruptivas, tenemos que situar la Inteligencia Artificial (IA), cuyo origen puede remontarse hasta el inicio de la digitalización, pero que en los últimos años ha mostrado un desarrollo de capacidades exponencial. A mediados del siglo XX, el padre de la digitalización Alan Turing (1950) ya reflexionaba sobre un posible método de aprendizaje de los programas computacionales a través de un sistema que incluyera mutaciones, por el cual el sistema pudiera recibir retroalimentación y, de esta manera, dirigir su aprendizaje en la dirección deseada. Desde entonces, diferentes hitos de desarrollo tecnológico han ido posibilitando nuevas funcionalidades de la IA, como las dependencias de largo plazo en las redes neuronales (Schidhuber y

Hochreiter 1997), las Redes Neuronales Convolucionales (Krishezvsky *et al*. 2012), Redes Generativas Antagónicas (Goodfellow *et al*., 2014), los transformadores (Vaswani *et al*. 2017), entre muchas otras.

Puesto que por el momento no hay una definición consensuada de la IA, procederé a partir de una definición propia. La IA son sistemas que sin la intervención activa de ningún ente biológico (humano, para lo que nos ocupa), toman decisiones adaptadas a un ambiente cambiante o a la experiencia. Sobre esta definición subrayo dos aspectos. El primero es que la IA no solo refiere a los modelos algorítmicos, puesto que su desempeño óptimo depende de un sistema que incluye, además de los propios algoritmos, un entorno de interacción, datos, capacidad computacional, etc. La valoración adecuada de la IA ha de incluir todos estos elementos. El segundo es la adaptabilidad de los sistemas de IA que, por lo tanto, imita una capacidad de la naturaleza. La IA cambia en respuesta al ambiente y a la evaluación de su comportamiento previo. Esto dificulta su análisis, puesto que no es un mecanismo fijo, y nos exige la consideración de sus características y capacidades generales. De todas maneras, es posible que este atributo dificulte el alcance del presente análisis prospectivo. La IA nos obliga no solo a mirar al futuro, sino a reevaluar nuestras prescripciones en función de su continua y acelerada transformación.

Con base en la definición anterior, dentro de la IA podemos considerar una variedad de sistemas, como: redes neuronales artificiales, aprendizaje automático (*machine learning*), aprendizaje profundo (*deep learning*), sistemas expertos, redes antagónicas, modelos de gran lenguaje, visión por computadora, entre otros. Esta variedad de sistemas, con diferentes utilidades y aplicaciones en diferentes sectores económicos, están mostrando un gran potencial de disrupción, con notables mejoras en cuestiones de productividad y efectividad en la sustitución de tareas que hasta ahora eran realizadas por seres humanos que requería, además, capacidades cognitivas medias e incluso altas. La funcionalidad de las tecnologías, como se apunta a continuación, está además relacionada con la disponibilidad de una gran cantidad de datos, necesaria para el entrenamiento de los modelos, y un gran poder computacional, necesario para el entrenamiento de los sistemas y su operación.

A continuación, se exploran algunas dimensiones críticas que deben considerarse dentro de una política digital de IA en sus dimensiones económicas. Como se argüirá, la complejidad de los efectos económicos de esta tecnología disruptiva requiere de una participación activa y de liderazgo de los Estados. Este planteamiento es, en consecuencia, contrario a la situación actual, de acuerdo a la cual el Estado ha dejado muchas áreas sin acciones preventivas,

frente a lo cual las grandes empresas tecnológicas (en adelante, *Big Tech*) han moldeado *de facto* el desarrollo de la IA de acuerdo a sus necesidades. El tema es de la mayor relevancia como pare que el futuro de estas tecnologías disruptivas y su consecuente impacto socioeconómico sea relegado a unos pocos e interesados actores, algunos de los cuales refieren incluso a la intervención regulatoria estatal como barreras al progreso tecnológico. Sobran experiencias sobre los peligros de la autorregulación y de las deficiencias de una falta de involucramiento activo por parte del Estado en el desarrollo de nuevas tecnologías. Por ello, a continuación consideraremos políticas públicas activas en una selección de temas económicos en los que el impacto de la IA requiere una atención prioritaria, como: concentración de mercados; impacto laboral; desigualdades en materia de desarrollo económico; y nuevos insumos estratégicos.

II. EFECTOS DE LA IA EN EL FUTURO LABORAL

Una de las afectaciones más importantes de la IA es su relación con la provisión de trabajo. En función de la clasificación de Korinek (2024) de esquemas de producción en función de la limitación y la capacidad de reproducción de factores productivos, la IA estaría rompiendo con el modelo productivo industrial en el que había un limitante natural a la reproducción del trabajo y una complementariedad de este con el capital. Por el contrario, la IA, como nueva forma de capital (tangible e intangible) tiene una mayor capacidad de sustituibilidad con el trabajo. Aunque se disponga de algunos factores por los cuales está tendencia pueda ser ralentizada o no sea total, frente a una potencia sustitución total del trabajo por IA de tipo general o suficientemente avanzada, los efectos son considerables para discutir la adopción de política económica preventiva y en respuesta. La menor demanda de trabajo implicaría un impulso hacia la disminución de los salarios, con la consecuente merma de ingresos de los trabajadores (en general, de la principal fuente de ingresos para la mayoría de la población) y de sus consecuencias socioeconómicas.

Centrándonos en los efectos de sustitución, Acemoglu y Restrepo (2018) explican que en aquellas actividades que puedan ser sustituidas por IA, la mejora de la productividad inducirá su incorporación, lo que resultará en un incremento de la producción superior a la de los salarios. A pesar del incremento de capital por trabajador resultante de una mayor productividad por la incorporación de procesos automáticos, la participación del trabajo sobre el producto total

se reduciría. La demanda creada a partir de esta mejora de la productividad debiera redundar en un incremento de bienes y servicios intensivos en trabajo. El incremento de demanda en estos sectores redundaría en un efecto contrario, de subida salarial, que paliaría parcialmente el primero. Sin embargo, este efecto dependería de la capacidad de absorción de estos sectores, existentes y los que se pudieran crear. Sobre la capacidad de absorción de trabajo de los sectores no afectados por la IA sería necesario valorar también si puede haber cambio de preferencias que redunden en una mayor valoración del trabajo humano, del papel que pueda generar el Estado en impulsar la demanda de estos bienes y servicios, así como el entorno institucional, con efectos en la determinación de los salarios y en la redistribución.

Las estimaciones actuales dirigen la atención a un importante grupo de actividades para las que se requieren capacidades cognitivas de nivel medio-alto, que van a ser afectadas por la IA, en particular por los Grandes Modelos de Lenguaje (GML). De acuerdo con las estimaciones de Eloundou *et al*. (2023), un 19% de las profesiones en EE. UU. verán afectadas por la IA más de la mitad de sus tareas, mientras que un 80% de las profesiones verán afectadas al menos un 10% de las tareas. Y sobre este conjunto de profesiones, las tareas cognitivas relacionadas con las tecnologías digitales no son ajenas. Por ejemplo, la programación se enriquece de herramientas como GitHub Copilot que al realizar tareas de programación más simplificadas reduce notablemente la necesidad de contratación de programadores; de hecho, las ocupaciones con mayores salarios tienden a estar más expuestas.

La robotización del trabajo tiene un sentido que va más allá de la sustitución y tiene una relación cercana con la IA. El origen de la palabra, que puede traducirse como trabajo forzado, servidumbre o esclavo surge de la obra de teatro de ciencia ficción de Karel Čapek, R.U.R. (Robots Universales Rossum) y da una nueva dimensión a lo que entendemos como robotización y una aplicación efectiva ya en el presente. De la misma forma que la mecanización del trabajo durante la industrialización se materializó en el control de la mano de obra en la línea de montaje como uno de los elementos característicos del taylorismo, la hiper vigilancia a través de herramientas de inteligencia artificial de los trabajadores puede concebirse como un "taylorismo con esteroides".

En la misma línea que el taylorismo había sido concebido como una racionalización del proceso productivo en pos de la eficiencia, las herramientas de inteligencia artificial subyugan al trabajador al cumplimiento estricto, bajo una vigilancia continua, de los objetivos establecidos, ya sean los tiempos de entrega, la satisfacción del cliente en cada una de sus interacciones merced

a los sistemas de puntuación, o la interacción del trabajador con el equipo informático.

La vigilancia de los trabajadores en su entorno laboral puede alcanzar límites distópicos que, en cierta medida, ya están siendo enfrentados por la regulación. La Ley Europea de AI prohíbe, por ejemplo, el uso de tecnologías de reconocimiento de emociones en entornos laborales. A pesar de las dudas sobre la cientificidad en la determinación universal de los gestos faciales como representativos de las emociones, ya hay herramientas de este tipo desarrolladas por prestigiosas universidades como el MIT o la Universidad de California y empresas como Human o HireVue que las emplean en el contexto de las entrevistas laborales (véase Crawford 2021). Los propios sesgos en el entrenamiento de los modelos de IA abren la posibilidad a sesgos por subgrupos en la identificación de dichas emociones, ya sea por cuestiones culturales o de género, entre otras posibles. El trabajador, más allá de tener un comportamiento correcto o adecuado, debe ser agradable y tener una expresión facial reconocible a tal efecto por una IA. La reedición de Tiempos modernos del siglo XXI podría ofrecer a un Charles Chaplin que, además de estar alienado por la definición de tiempos y minimización de tareas en una línea de producción optimizada por IA, sonriera de forma convincente para que la IA valorara positivamente su aportación al óptimo desarrollo del ambiente laboral. Por otro lado, cabe una consideración ética sobre si las emociones constituyen un elemento privado sobre el que no deben establecerse criterios de medición de rendimiento.

Mientras que la IA puede ser una herramienta importante para medir mejor la productividad, el uso de estas medidas puede ir en varias direcciones. Pueden ser los propios sindicatos quienes acumulen mejores datos sobre el avance en la productividad y, de esta manera, dotarse de mejores herramientas para la negociación colectiva. Sin embargo, podemos suponer una capacidad asimétrica en el entorno laboral a favor del patrón para el uso de estas herramientas. Una medición exhaustiva de la productividad puede individualizar totalmente la caracterización de un trabajador, cuestión que puede afectar la misma esencia de la negociación colectiva.

Además, habría que evitar prácticas como: la recolección de datos fuera del horario laboral o del centro de trabajo; el uso de datos que puedan implicar una discriminación; afectaciones a la privacidad de los trabajadores; o el uso de la IA para prevenir acciones en defensa de los derechos de los trabajadores. Por ejemplo, una recolección de datos intensa puede ayudar a situar trabajadores más activos en la defensa de sus derechos o para la justificación de despidos

de trabajadores involucrados en estas actividades. Debido precisamente a la asimetría inherente a las relaciones laborales, el consentimiento puede no ser un criterio válido para valorar la vigilancia en el lugar de trabajo.

Los nuevos modelos económicos facilitados por las plataformas ofrecen nuevos esquemas laborales, frente a los cuales puede ser necesario reformular la relación laboral y la definición de trabajo. Tradicionalmente, se ha considerado la propiedad del capital como un elemento distinto de empresariado, frente al que se define al trabajador como proveedor de la mano de obra. Sin embargo, es preciso atender si la provisión (de una parte) del capital continúa siendo un elemento para determinar la asimetría en la relación empresa-trabajadores. En la actualidad, podemos tener esquemas de contratación (esquemas de asociados en diferentes plataformas de transporte y reparto) en el que hay una relación de dependencia y asimetría, pero la plataforma se emplea como una herramienta para la redefinición de la relación laboral.

III. INCREMENTO DE CONCENTRACIÓN A PARTIR DE LA DIGITALIZACIÓN Y DE LA IA

El cambio de modelo hacia una economía más digitalizada con un mayor uso de IA, exige una reformulación de aspectos que afectan a la política de ordenamiento industrial, con especial atención a los efectos en la competencia. En este nuevo ordenamiento, es preciso referir al *Big Tech*, como se conocen a las grandes empresas tecnológicas que pueden hacer uso de su posición de dominancia para su beneficio. Como veremos brevemente, hay razones intrínsecas a la industria digital y su ordenamiento para atender a la concentración que puede darse en el sector, que se puede intensificar con la introducción de la IA. El origen nacional de este control empresarial no es baladí y está relacionado con los diferentes modelos de regulación (o la ausencia de ella) que se plantean. Dentro del *Big Tech*, se destaca el *Big Five*, como las cinco empresas tecnológicas más importantes (Alphabet Google-, Amazon, Apple, Meta y Microsot) todas ellas estadounidenses. Aunque fuera de este grupo, no podemos evitar la consideración de empresas como Huawei, Alibaba, Tencent o ByteDance de China.

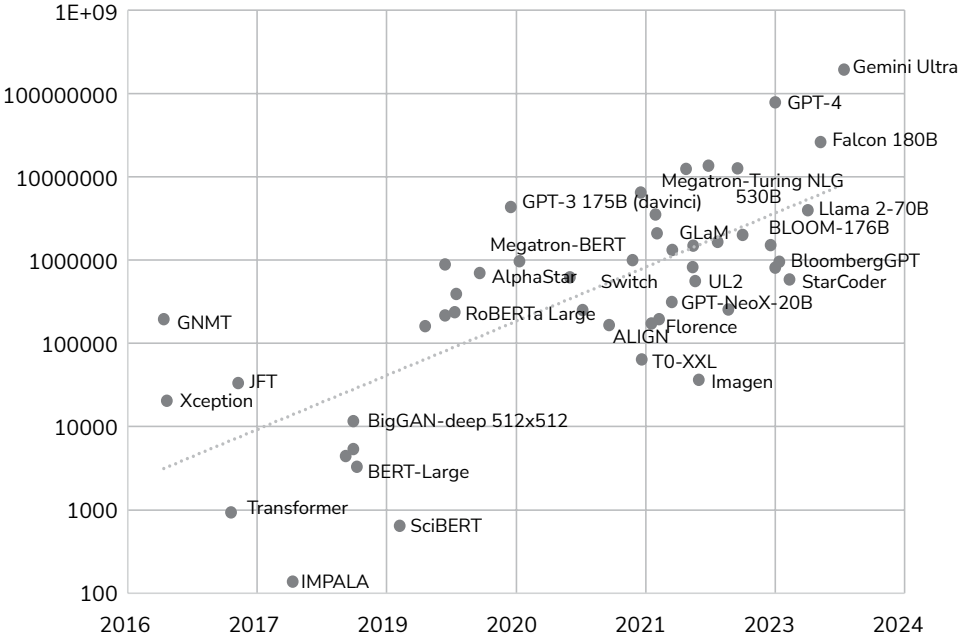
Varias de estas empresas operan bajo sistemas de plataforma, ya sea controlando directamente una red social o el sistema operativo, y ejercen un control directo sobre el acceso de millones de usuarios a otras funcionalidades que pueden ser provistas por empresas competidoras. En ese sentido, su participación en un sistema económico puede ser la de una infraestructura de

servicios públicos (*public utility*) y su ordenamiento puede seguir principios similares, en tanto que pueden compartir características de monopolio natural (véase Tirole 2024). De hecho esta situación de monopolio natural se presenta en los sectores digitalizados de una forma más amenazante que la de otros monopolios naturales en el pasado, porque mientras que estos tendían a estar limitados nacionalmente, la facilidad de las tecnologías digitales para operar internacionalmente, junto con las regulaciones de liberalización comercial en materia digital y las nuevas utilidades de IA que reducen el efecto de barreras naturales como el idioma favorecen que las *Big Tech* puedan extender su poder monopólico a nivel global.

Esta capacidad de concentración industrial se puede ver reforzada por la IA por, al menos, dos razones: la primera son los recursos que requieren los tres pilares de la IA, y la segunda refiere a la capacidad de desarrollo acumulativo que permite la IA. En relación con la primera, la IA requiere de tres elementos fundamentales para su desarrollo: capacidad computacional, algoritmos y datos. La mayor complejidad de los últimos Modelos de Gran Lenguaje (LLM, por sus siglas en inglés) requiere un mayor poder computacional para su ajuste. Aunque, el rango de estimación sobre el número de parámetros empleados es amplio, debido a que las compañías no revelan esta información, se estima que es superior al billón para los últimos modelos como ChatGPT-4 y Gemini Ultra, con poderes computacionales requeridos para su entrenamiento de 50 mil y 21 mil millones de Petaflops respectivamente. Esto no es un ejemplo aislado, sino la última realización de una tendencia exponencial en el costo de entrenamiento de los modelos más complejos de IA. En esta línea se estimaría que tan solo el poder computacional requerido para el entrenamiento de Gemini Ultra rozaría los 200 millones de dólares. Estas capacidades de procesamiento no están al alcance de pequeñas compañías, por lo que representan una importante barrera de entrada.

La disponibilidad de datos es esencial para el desarrollo de capacidades de los modelos de IA, en particular para su entrenamiento. Una de las ventajas que las Big Tech disfrutan sobre posibles competidores reside, precisamente, en el gran número de datos que tienen de sus usuarios a través de sus interacciones en sus plataformas y que pueden ir desde compras y consultas hasta simples (y muchas veces incontroladas) interacciones, como permanecer un tiempo mayor observando en pantalla determinada publicación. Cada uno de estos datos individuales puede no tener gran valor en sí mismo, pero de forma colectiva pueden materializar un valor superior al costo marginal de expandir a un individuo el uso de una plataforma, que es cercano a cero.

GRÁFICA 1 Costo de entrenamiento por uso computacional
(en US$, ajustado por inflación)



FUENTE: Elaboración propia con datos de HAI (2024)

Una de las fuentes de datos más masivas y baratas (o gratuitas) es la vinculada a publicaciones abiertas en Internet. Estos datos han sido utilizados por ejemplo para el entrenamiento de los Modelos de Gran Lenguaje (LLM, por sus siglas en inglés). Sobre esta materia existen grandes dudas debido a que las propias empresas desarrolladoras son reticentes a compartir la fuente de datos con las que se entrenan estos modelos, debido a las potenciales reclamaciones que pudiera haber por los derechos de esas publicaciones. No obstante, de las declaraciones de líderes en la industria refiriendo a que el uso de la información pública es lícito y de la necesidad de dicha información, podemos concluir que está siendo empleada. La disponibilidad de datos abiertos puede estar significando ya un límite al crecimiento exponencial de las habilidades de los modelos de IA, en relación a que estos modelos ya estarían utilizando para su entrenamiento toda la información pública disponible en Internet. Además, debido a que la propia participación de la IA en la creación de contenido es creciente, el porcentaje de "datos humanos" sería cada vez menor, lo que reduciría la cantidad de datos disponibles para entrenamiento. A este respecto,

se puede pensar que se esté alcanzando algún límite sobre los insumos de la IA y, por lo tanto, sobre el crecimiento de sus capacidades. En términos del límite de la disponibilidad de datos hay, al menos, dos fuentes que pueden evitar cuellos de botella en el desarrollo de modelos. La primera es la mayor captura de datos a través de dispositivos permanentemente conectados a la red, como los que se agrupan pago el Internet de las Cosas. Un mayor número de *wearables*, cámaras, micrófonos, etc. continuamente conectados gracias a las redes 5G pueden multiplicar la cantidad de datos disponibles, aunque estos no se encuentren de forma pública en la red. De hecho, esta posibilidad no hace más que incrementar el valor estratégico de las empresas por posicionar este tipo de herramientas en el mercado de la forma más masiva posible. La segunda son los datos sintéticos generados a partir de diferentes procesos como las redes antagónicas, los autocodificadores variacionales o las simulaciones. Por ejemplo, la empresa Waymo de conducción autónoma había acumulado 20 millones de kilómetros de manejo en el mundo real, había multiplicado por mil esta experiencia en conducción en simuladores (Waymo 2024). Las simulaciones pueden enriquecerse continuamente con datos provistos del mundo real, pero se pueden combinar de formas novedosas para testar la optimalidad de la respuesta algorítmica.

En relación con la ventaja acumulativa que puede tener la IA, esta se puede presentar por varias razones. Por un lado, las posiciones de poder de mercado obtenidas por las empresas dominantes pueden facilitarles el acceso a segmentos recién creados o, en general, con mayor potencial de crecimiento. Por ejemplo, las sinergias en la explotación de datos pueden facilitar que las empresas dominantes en sectores digitales extiendan su capacidad a otros sectores, como el financiero, en el que la capacidad de previsión puede constituir un diferencial en materia de productividad. Lo anterior plantea retos para propiciar una mayor competencia en el sector. Por ejemplo, si un uso continuado de determinado modelo de IA puede adaptarlo a los criterios y requerimientos específicos de un usuario, este cliente tiene menos incentivos para cambiar de empresa y de modelo de IA. ¿Puede considerarse algún tipo de portabilidad del historial o de los datos de entrenamiento individuales, sobre los que el usuario ejerza mayor control?

Por ello, la atención la vigilancia de la competencia debe superar los criterios tradicionales que se consideran en materia de afectación de las concentraciones, como el bienestar de los consumidores a través de los precios. Las empresas pueden, por ejemplo, acumular poder de mercado a través de la acumulación de datos. Se debe prestar una mayor atención a las estrategias

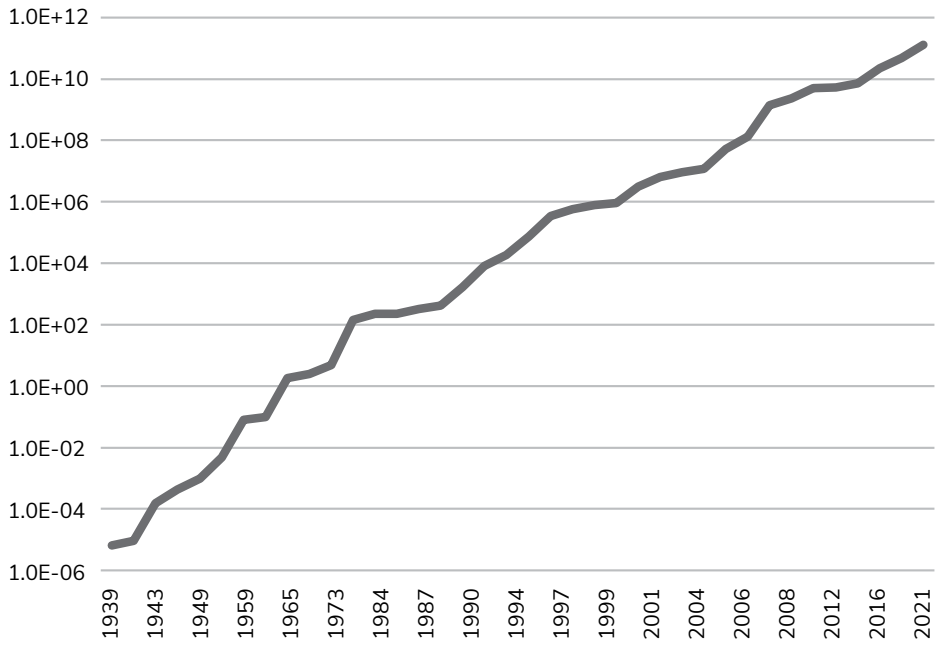
entre compañías, o subsidiarias de la misma compañía, en el intercambio de datos. El consentimiento en el uso de datos como localización, datos relativos a la salud, entre muchos otros, ligados al uso de una aplicación en concreta puede ser comercializado para otros usos no considerados por el usuario, por lo que los modelos ligados al consentimiento del usuario pueden resultar insuficientes. Las leyes de competencia y los organismos encargados de vigilar la competencia deben analizar los efectos de fusiones también en materia de datos.

Por último, es preciso mencionar cómo el propio desarrollo de IA puede acelerar nuevos desarrollos tecnológicos, lo que puede desencadenar una aceleración de desarrollo tecnológico concentrada. Aunque para algunos todavía pueda entenderse como ciencia ficción, los expertos en IA comentan de forma cada vez más generalizada sobre la posibilidad de que en el medio plazo se desarrolle inteligencia artificial general (IAG), con capacidad de desarrollo autónomo en todos o en una gran variedad de campos, o de que se alcance la singularidad tecnológica, por la cual la propia IA se desarrollaría a sí misma, generando un bucle de aprendizaje exponencial. Desde luego, es posible que no se estén considerando algunos factores esenciales para el correcto desarrollo de estas tecnologías y que las capacidades no respondan exactamente a las predicciones más fantasiosas. Sin embargo, no debemos obviar que la IA debe considerarse como una tecnología particular en tanto que puede dotar de los elementos para avanzar en el propio desarrollo tecnológico.

La afectación de una gran concentración de mercados en un segmento tan amplio como la economía digital y la IA, que están en pleno crecimiento y con profundas afectaciones a varios sectores socioeconómicos y de desarrollo científico y técnico son muy importantes. De igual manera que los instrumentos a partir de los cuales opera la concentración van más allá de los precios, la afectación al bienestar de los consumidores y ciudadanos se puede dar a través de otras dimensiones. Por ejemplo, estas empresas pueden usar su mayor poder a través de las redes para afectar a la misma regulación de la que son objeto. ¿Han de tener responsabilidades específicas las empresas de redes sociales que pueden afectar el curso de una contienda electoral? En otro orden, la negativa de las empresas para revelar aspectos sobre la arquitectura de los modelos logarítmicos empleados, el entrenamiento de dichos modelos y los tipos de datos empleados dificultan el desarrollo científico de la IA, así como una regulación precautoria. Debido al tamaño de los grandes jugadores en la economía digital y la IA y su afectación a aspectos de desarrollo tecnológico que incluyen, desde luego, tecnologías de vigilancia gubernamental y militares,

la actuación de los gobiernos debe considerarse en muchas ocasiones en connivencia con las principales empresas desarrolladoras de IA.

GRÁFICA 2 Computaciones por segundo por dólar (contante de 2023) en su mejor desempeño



FUENTE: Elaboración propia con datos de Kurzweil (2024, 165-166)

Como elemento contrario a las tendencias de concentración tenemos el exponencial aumento de la capacidad computacional por dólar constante. Esto indica que, para un mismo poder de compra, se puede alcanzar cada vez un mayor poder computacional, lo que abarata rápidamente el acceso a capacidades computacionales de alto desempeño. Esto no necesariamente implica una reducción de las barreras de entrada en materia de competencia, puesto que la competencia puede estar dándose en niveles de poder de computación que siguen dejando fuera a un gran número de compañías, por la complejidad de los algoritmos empleados y el alto número de datos requeridos para su entrenamiento. Sin embargo, puede suponer una reducción de la brecha de alcance de los países en desarrollo en pocos años de poderes computacionales a precios reducidos.

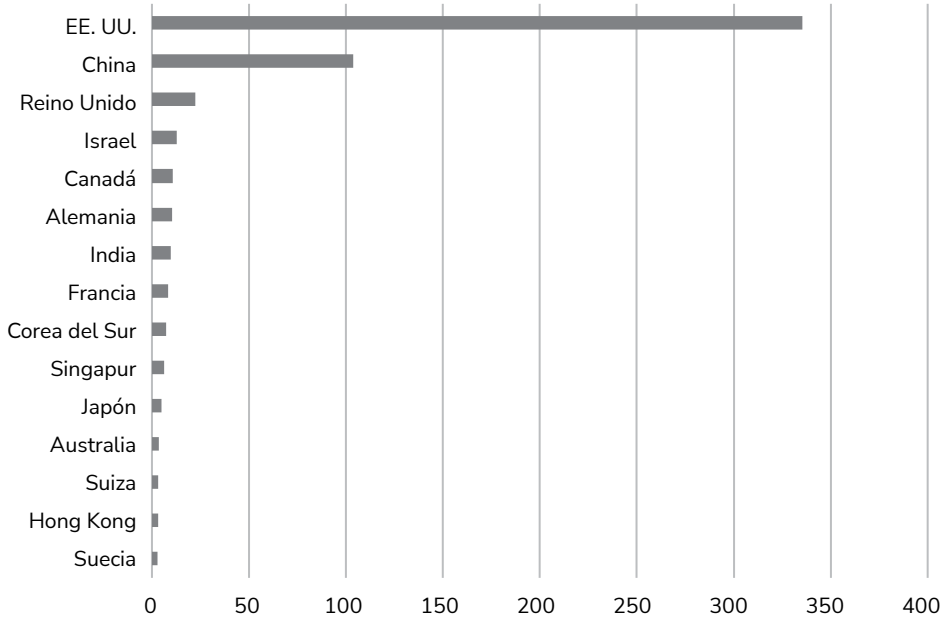
IV. ACUMULACIÓN DE CAPACIDADES DE DESARROLLO TECNOLÓGICO Y AFECTACIÓN AL PROCESO DE DESARROLLO

Varios de los elementos que hemos mencionado pueden, de forma conjunta, limitar notablemente la capacidad de *catch up* tecnológico de los países en desarrollo y, de esta forma, a su capacidad de crecimiento económico y desarrollo socioeconómico. La capacidad de desarrollo tecnológico tiene una importante dimensión acumulativa y la IA puede reforzar este aspecto. En la medida en que los nuevos conocimientos e infraestructura requieren de los anteriores, se dificulta desarrollar uno sin los primeros. Estos no solo dependen del conocimiento, sino que también involucran a insumos estratégicos, know how, experiencia institucional. Por ejemplo, es complejo tener una capacidad de infraestructura digital si no se dispone de infraestructura energética a la par, por el alto consumo energético de estas instalaciones, además de los requerimientos que puedan tener las empresas en términos de generación verde.

A no ser que las nuevas tecnologías se desarrollen en campos completamente diferenciados y no complementarios, o que los costos relacionados con desarrollos previos se reduzcan drásticamente, un desarrollo tecnológico acelerado como el que se puede dar bajo la IA puede aumentar notablemente las brechas entre países y dificultar su corrección. Algunos efectos de esta restricción de acceso tecnológico pueden ser beneficiosos, puesto que se puede ejercer un mayor control en el uso de tecnologías con demasiada capacidad y con potenciales efectos perjudiciales. Por ejemplo, mantener cierto control sobre modelos de IA muy potentes y con capacidad para crear nuevos agentes biológicos infecciosos, puede evitar un amplio acceso a esta capacidad tecnológica y limitar que esta sea accesible por, por ejemplo, grupos terroristas. (capaces potencialmente de nuevos desarrollos biológicos). Sin embargo, a su vez el limitado acceso tecnológico impide nuevos desarrollos con beneficios civiles.

De hecho, como se ha puntualizado anteriormente, los nuevos modelos de IA con mayores capacidades, pero, a su vez, mayores requerimientos de poder computacional y datos, requieren mayores inversiones, efecto por el cual se tiende a retroalimentar la concentración corporativa. En esa misma dirección, la concentración de las grandes empresas en EE. UU. (principalmente), en China (en segundo lugar) y en otros países desarrollados, se replica la concentración

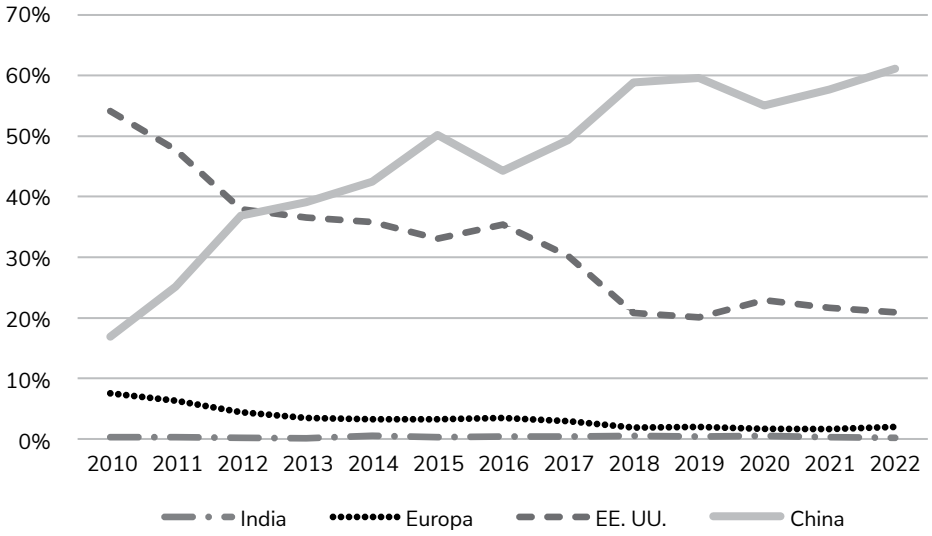
GRÁFICA 3 Inversión corporativa acumulada 2013-2023 (en modelos de IA mil millones US$)



FUENTE: HAI (2024)

de empresas e inversión privada a nivel nacional. De esta forma, durante los diez últimos años la inversión privada se ha concentrado de forma muy notable en EE. UU. con más de 300 mil millones de inversión privada en modelos de IA. Le sigue China, con menos de un tercio de la inversión privada, aunque con resultados interesantes en relación con la inversión realizada. Después de los dos hegemones a nivel mundial, tenemos una mayor dispersión entre países desarrollados. Europa, como región y la UE como actor político colectivo más importante, puede aglutinar una tercera alternativa, si bien con un peso acumulado muy alejado en materia de inversión privada. No podemos dejar de mencionar, que el peso de los países en desarrollo es mínimo, con la presencia de India como actor destacado, de donde cabe retomar la iniciativa de Infraestructura Pública Digital (DPI, por sus siglas en inglés), como modelo alternativo de desarrollo digital (incluyendo la IA) con una atención particular a los retos diferenciados que experimenta el Sur Global.

GRÁFICA 4. Patentes de IA concedidas como porcentaje mundial por áreas geográficas



FUENTE: HAI (2024)

La concentración geográfica de las patentes específicas en IA permite realizar algunos ejercicios de prospectiva sobre cómo se puede comportar la concentración tecnológica en el futuro. En términos agregados se observan claramente dos fenómenos. En primer lugar, el desarrollo tecnológico de IA se encuentra enormemente concentrado a nivel país y esa concentración, lejos de mitigarse, se ha incluso acrecentado durante los últimos años. Mientras que EE. UU. y China concentraban en torno al 71% de las patentes en IA en 2010, ese porcentaje conjunto supera el 82% en 2022. De hecho, en esa tendencia una de las regiones que más dinamismo ha perdido ha sido Europa que, mientras que en 2010 contaba con el 7.54% de las patentes en IA ese porcentaje apenas supera el 2% en 2022. Es cierto que estos números relativos deben tenerse en cuenta con respecto a un número exponencialmente creciente de patentes en IA durante los últimos años, pero la pérdida de dinamismo relativo es muy significativa. Por un lado, el sólido avance regulatorio de la UE (Ley de Servicios Digitales, Ley de Mercados Digitales, Ley de Inteligencia Artificial) ha establecido una interesante estructura regulatoria de competencia asimétrica que limita el control de las *Big Tech*, además que establece criterios de defensa de los derechos humanos en los desarrollos de IA. Sin embargo, no podemos obviar que el desarrollo tecnológico en IA está precisamente protagonizado por las grandes empresas

y que una regulación más estricta en Europa puede reducir los incentivos a invertir en la región.

En segundo lugar, es llamativa la sustitución en el primer lugar de desarrollo de patentes de IA de EE. UU. por China. Mientras que en 2010, EE. UU. lideraba la concesión de patentes en IA con un 54.13% y China se situaba en un 16.89%, los papeles se han tornado y en 2022 el país norteamericano se sitúa en un muy reducido 20.90%, mientras que China ya alcanza el 61.13%. Desde luego que el número absoluto de patentes no ofrece toda la imagen de desarrollo tecnológico y que estas pueden tener un valor económico o centralidad distintas, pero refuerza una tendencia de sustitución en el liderazgo tecnológico que se viene presentado en diferentes segmentos económicos. Si, por ejemplo, apuntalamos el dato con las patentes por habitante, EE. UU. sí se encontraría por delante de China, con 4.23 patentes por cada 100,000 habitantes, frente a las 2.51 del gigante asiático. Sin embargo, también en este rubro EE. UU. perdería el liderazgo internacional, puesto que Corea del Sur contaba en 2022 con 10.26 patentes por cada 100,000 habitantes en IA. De esta manera, lo que advierte el desarrollo tecnológico es que es muy posible que la IA se mantenga como un sector tecnológico y económico altamente concentrado, pero que este vaya pivotando cada vez más hacia Asia del Este.

Otras regiones del Sur Global se mantienen olvidadas en el desarrollo tecnológico de IA. El país con mayor presencia es India, que apenas ha superado el medio punto porcentual sobre el total de patentes de IA en algún año y que en 2022 se situaba en el 0.23%. Los países de Latinoamérica, Sudeste Asiático y África ofrecen una baja representación en desarrollo de patentes en esta tecnología disruptiva, muy inferior desde luego a su peso económico, lo que ofrece una perspectiva negativa sobre el efecto de la desigualdad regional en materia de economía digital durante los próximos años. Como elemento esperanzador en la dirección contraria se sitúa el crecimiento de modelos base abiertos durante los últimos años. Por ejemplo, de los 149 modelos base que se publicaron en 2023, casi dos tercios eran de código abierto, mientras que la proporción era la inversa apenas dos años antes. La mayor expansión de modelos base de código abierto puede ofrecer espacios de colaboración para los países del Sur Global y un desarrollo tecnológico más inclusivo.

V. IMPULSO A LA INTELIGENCIA ARTIFICIAL Y SUSTENTABILIDAD

La implementación de la IA puede tener impacto en uno de los objetivos esenciales de la política económica a nivel global, la mitigación del cambio climático, particularmente a partir de la limitación y potencial reducción de los gases de efecto invernadero. La IA tiene efectos medioambientales negativos y beneficiosos y sería necesario mantener una investigación actualizada sobre su resultado neto.

Por un lado, se requieren grandes cantidades de energía para el funcionamiento de la infraestructura computacional de los centros de datos y centros de procesamiento de los grandes modelos de IA. De esta forma, frente a sus ventajas en materia de seguridad y eficacia, se debe tener en cuenta que la computación en la nube es muy intensiva en recursos energéticos y, en particular, en electricidad. A pesar de los esfuerzos de la industria por aumentar el empleo de energías renovables, la huella de carbono de la industria computacional ha sido comparada con la de la industria de aviación. Por tomar como referencia, Luccioni *et al.* (2022) han estimado la huella de carbono de BLOOM, un modelo de lenguaje de 176 mil millones de parámetros. Tan solo el entrenamiento de BLOOM emitió 24.7 toneladas CO2 equivalentes, bajo la consideración del empleo del poder computacional, cuantía que se duplicaría teniendo en cuenta los procesos de producción del equipo y otros procesos operacionales. Bajo sus propias estimaciones estas emisiones son bajas en comparación a las necesarias en otros modelos. GPT-3, con un número de parámetros similar, habrían generado diez veces más emisiones (502 toneladas de CO2 equivalentes). Es cierto que la mejora en la eficiencia del entrenamiento de los modelos ha mejorado, pero, a su vez, están aumentando el número de los parámetros de los modelos y, en consecuencia, el poder computacional necesario para su operación. Para 2020, Freitag *et al.* (2021) estimaban una participación de las TIC (tecnologías de la información y la comunicación) de entre un 2.1% y un 3.9% en las emisiones de efecto invernadero. Además del amplio rango de la estimación, que responde a la baja transparencia en esta materia de las empresas tecnológicas, debemos tener en cuenta que esta participación puede aumentar durante los próximos años. Este aumento no solo estaría relacionado con un uso más extensivo de las herramientas digitales, sino porque las herramientas de IA son más intensivas en uso energético. Por ejemplo, mientras que una búsqueda generaría 0.2g de CO2 equivalente, una búsqueda por ChatGPT podría multiplicar entre 50 y 90 veces ese impacto medioambiental (Vanderbauwhede 2023), por lo que una

sustitución de los métodos de búsqueda por un uso más intensivo de IA podría tener notables efectos en las emisiones totales.

El uso y entrenamiento de la IA también es intensiva en agua, un recurso natural sobre el que hay una presión cada vez mayor. Algunas estimaciones apuntan a que la demanda global de IA podría ser responsable en los próximos años a un consumo equivalente al de la mitad del Reino Unido (Li *et al*. 2023). Además de la operación de la infraestructura computacional, los elementos computacionales son intensivos en el uso de materiales críticos. Por un lado, los componentes informáticos pueden implicar un uso de 400 veces en peso de materias primas frente al producto acabado utilizado. Muchos de estos materiales son minerales raros, cuya extracción está relacionada con métodos contaminantes y grandes desplazamientos de tierras, debido a la baja concentración de estos materiales incluso en minas especializadas. Además, los componentes electrónicos como deshechos pueden tener efectos medioambientales notables, debido a la presencia de elementos muy contaminantes, como el mercurio.

Para mitigar el impacto medioambiental de la IA, es necesario prestar atención a las regulaciones y políticas de control de impacto ambiental considerados. Por ejemplo, la Ley de Inteligencia Artificial establece pautas de favorecimiento a modelos de IA energéticamente eficientes (Parlamento Europeo 2024). De esta forma, entre la regulación técnica que deben presentar los proveedores de modelos de IA general se encuentra el gasto energético de procesamiento, así como el empleado para el entrenamiento del modelo. Además, se incluye el criterio de protección medioambiental y su monitoreo en la implementación de los modelos de IA. En EE. UU., aunque se presentó a nivel federal la Ley de Impacto Medioambiental de la IA en febrero de 2024, a la fecha de realización de este capítulo, la propuesta estaba siendo estudiada por el Senado.

Frente a estos efectos negativos, se deben considerar, a su vez, los potenciales efectos positivos de la IA. En términos de lo que se ha discutido, este conjunto de mejoras tecnológicas tiene la capacidad de sustituir un trabajo equivalente que, de realizarse por humanos, podría implicar un consumo de recursos superior. Es decir, en términos del producto ejecutado, queda abierta la discusión sobre la eficiencia de la IA. Por otro lado, se debe tener en cuenta que su uso intensificado tendrá implicaciones medioambientales y que la comparación no debe limitarse a una eficiencia relativa.

VI. ELEMENTOS PARA UNA AGENDA ECONÓMICA EN MATERIA DE IA

La disrupción que significa la IA como elemento particular y como parte de una nueva economía digital afectará de forma innegable e inevitable varios aspectos económicos y del régimen laboral de todos los países. Ante cambios de esta envergadura, es necesario mantener una agenda prospectiva. A continuación, se discuten brevemente algunos aspectos que debe tratar dicha agenda, con especial atención a la cuestión laboral.

El desarrollo tecnológico que implica la Revolución Digital en general y la IA en particular ofrece enormes oportunidades de mejora en la productividad. Sin embargo, la rentabilidad de algunos modelos de negocio vinculados a las plataformas digitales y de uso de la IA puede estar vinculado a prácticas indeseadas, como la evasión del régimen laboral mediante la simulación a través de contratos de asociados, sistemas tayloristas de vigilancia exhaustiva o, de forma relacionada, la total individualización del trabajador como estrategia velada de debilitamiento de la negociación colectiva. Las instituciones, formales e informales, de un régimen laboral son esenciales para garantizar derechos laborales y niveles salariales dignos. Estas condiciones se constituyen, a su vez, como elementos centrales para el mantenimiento o la consecución de una sociedad democrática estable.

La política regulatoria (de forma general, así como la acción legislativa, de manera particular) en materia laboral y las instituciones de vigilancia de su correcta aplicación han de estar continuamente actualizadas sobre las formas de denegación de derechos laborales que puede suponer la IA y los nuevos modelos digitales. A este respecto, se debe evitar la asunción de lemas promulgados por el Big Tech sobre modelos de trabajo flexible y buscar la forma actualizada de aplicar principios esenciales sobre el trabajo, acordados internacionalmente; por referir a un principio fundamental, como se establece en preámbulo de la constitución de la Organización Internacional del Trabajo: "el trabajo no es una mercancía". Evidentemente, el trabajo no corresponde a esta categoría, pero de forma añadida se debe entender que "el trabajo no debe tratarse como una mercancía". Bajo esta concepción, no es exagerada la aplicación de regulación precautoria en la materia y sospechar escépticamente los instrumentos de evaluación empleados en los sandboxes regulatorios debido, entre otras cuestiones, al poder político de las Big Tech.

El segundo y posiblemente más temido efecto de la IA en materia laboral es el desempleo tecnológico, por enunciarlo en los mismos términos en los que lo hizo

J. M. Keynes. Frente a esta posibilidad que se puede materializar en diferentes formas, debemos evitar argumentos inductivos sin sustento. Que frente a anteriores cambios tecnológicos hayan surgido nuevos sectores y posibilidades de empleo no implica que esto deba ser necesariamente cierto ante una transición productiva como la de la Revolución Digital. Las realizaciones pasadas solo pueden emplearse en tanto las condiciones se mantengan y las condiciones económicas están cambiando radicalmente. Descartada la negación, la inacción es muy costosa. Hasta el más férreo defensor de las fuerzas de mercado entenderá que el periodo de transición hacia un nuevo equilibrio en el mercado laboral es mucho más lento y socialmente costoso que en otros mercados. Un planteamiento que abogara por el parón del desarrollo de IA, aunque pueda ser exitoso en sectores muy específicos, es inmanejable a escala general. La búsqueda por la mejora de la competitividad arrastrará irremediablemente a una adopción intensiva de la tecnología. El desempleo tecnológico no refiere únicamente a la pérdida de trabajos, sino a una reestructura de la fuerza laboral con un posible aumento de la desigualdad.

El primer tipo de medidas que se puede dirigir a la sustitución de mano de obra es a su mitigación. Esto no necesariamente debe implicar que los esfuerzos se dirijan a parar el desarrollo tecnológico, cuestión que, como hemos señalado es difícil de manejar debido a que no está alineado a los incentivos de la carrera por el desarrollo de la IA. Sin embargo, el mantenimiento de control humano en sistemas y tecnologías puede responder a cuestiones éticas como la no delegación de agencia a sistemas automáticos. De forma similar a como las respectivas agencias reguladoras de la aviación a nivel internacional y nacional obligan a la presencia de dos pilotos en los vuelos comerciales, la regulación puede limitar la sustitución de la agencia humana en actividades como la conducción, el cuidado de personas o el diagnóstico, entre otras. Este no evita que se incluyan y desarrollen tecnologías de IA, sino que la sustitución sea absoluta.

En muchas actividades el mantenimiento de la agencia humana puede significar una notable reducción de empleo en respuesta al incremento de la productividad. Por ejemplo, el uso de la IA en actividades de programación ha reducido notablemente el número de horas-hombre requeridas para desarrollos en materia de software. Esto es, la IA está sustituyendo, aunque sea parcialmente, un importante número de empleos de tipo cualificado, cuyas actividades se concentran en un menor número de personas con altas capacidades para, en complementación con la IA, mejorar ostensiblemente la productividad. Esta concentración de actividades productivas cualificadas en un reducido número

de personas puede generar un notable aumento de la desigualdad, como ya lo han observado Acemoglu y Restrepo (2022) para actividades automatizadas. Frente a ello, cobran cada día más relevancia propuestas como la del Ingreso Básico Universal (IBU).

A través de un Ingreso Básico Universal, el Estado aseguraría un ingreso mínimo a todos los ciudadanos. Aunque se ha concebido como universal para que no existan desincentivos teóricos para la realización de actividades productivas, la universalización implicaría un costo inasumible incluso para los países desarrollados. Es por ello que, en los experimentos aplicados de IBU, como los realizados en Canadá y en Finlandia, se han dirigido a personas desempleadas y familias sin ingresos. Como se ha encontrado en el caso de Finlandia, el IBU tiene un efecto significativo en la reducción del estrés vinculado al desempleo y la falta de ingresos, sin efectos negativos en la empleabilidad (*véase* Kangas *et al*. 2021). No obstante, el principal impedimento para una aplicación generalizada del IBU, la enorme carga fiscal, afecta diferencialmente a los países en desarrollo, con capacidades fiscales mucho más reducidas.

La aplicación de una medida de IBU requiere de recursos fiscales adicionales. Frente a ello, se ha planteado que el origen de estos recursos sea, precisamente, la creación de impuestos específicos sobre tecnologías de IA. La noción de la propuesta es sencilla, que un impuesto sobre la tecnología que está desplazando empleos internalice este costo. Sin embargo, un impuesto de este tipo se enfrenta con algunos retos. En primer lugar, sería necesaria una definición estricta de cuáles son las tecnologías comprendidas dentro de la IA a las que se aplicaría un impuesto específico, cuya justificación frente a otras tecnologías que también mejoran la productividad y consecuentemente reducen el número de horas de trabajo requeridas es complejo. En segundo lugar, no existe una equivalencia entre las ubicaciones de las tecnologías de IA que mejoran la productividad y los empleos afectados. A nivel internacional, la concentración de las actividades de IA en países desarrollados propiciaría que estos países concentraran las capacidades fiscales derivadas de estos nuevos impuestos. Por último, la creación de impuestos específicos a nivel nacional sobre tecnologías de IA crearía una desventaja competitiva para los países que los promulgaran, lo que desincentiva acciones individuales a nivel país en esta materia. Estos retos no son insalvables, pero apuntan hacia una mayor coordinación a nivel internacional.

Como se ha señalado, varias de las características de la economía digital apuntan hacia una mayor concentración de los mercados y la revolución de la IA exacerba esta presión. Frente a esta tendencia propia de las características

de la industria digital y de la IA, es preciso establecer una vigilancia específica al sector. De hecho, los países en desarrollo pueden no tener suficiente fuerza regulatoria para enfrentar a estos gigantes tecnológicos globales. Ya se ha advertido la presión ejercida a entidades político-económicas más poderosas, como la UE, o la guerra política a nivel internacional de la plataforma política X frente al poder judicial de Brasil. Por ello, las instancias de coordinación internacional para reducir la presión de las grandes tecnológicas son esenciales. Como se ha subrayado, esta política de vigilancia debe superar las dimensiones clásicas de afectación al bienestar de los consumidores e incluir otros elementos que, aunque más complejos, pueden tener una mayor relevancia en el medio y el largo plazo, como la influencia política e institucional, o el control sobre los datos.

En materia de mitigación medioambiental, es preciso adoptar una regulación precautoria. Por un lado, de forma similar a la regulación europea, la mayor transparencia exigiría que como parte de la descripción técnica de los modelos de IA, se exija la publicación de su uso energético, tanto el puntual como el empleado durante el entrenamiento. En función de lo anterior, la política pública puede favorecer criterios de contratación en las compras gubernamentales en favor de una mayor eficiencia energética. Por otro lado, como parte de la política energética nacional se debe favorecer una instalación de capacidad de generación energética sustentable en aquellas regiones en que se quiera favorecer la atracción de inversión en centros de datos.

En términos regulatorios de la IA, nos encontramos en la actualidad ante tres modelos efectivos. En primer lugar, tenemos el modelo europeo, liderado por la regulación propuesta desde la Unión Europea, caracterizados por la Ley de Servicios Digitales (Diario Oficial de la Unión Europea 2022a), la Ley de Mercados Digitales (Diario Oficial de la Unión Europea 2022b) y la Ley de Inteligencia Artificial (Parlamento Europeo 2024). Frente a otros modelos, observamos que es una regulación con importantes elementos frente a la identificación de una competencia asimétrica, con exigencias a los identificados como *gatekeepers*. En particular, en materia de IA, se prohíben o limitan usos en función de que puedan afectar a los derechos humanos. Por ejemplo, se prohíbe el uso de herramientas de identificación emocional en el trabajo o la identificación facial de forma continua como parte de modelos de seguridad nacional en espacios públicos. La regulación asimétrica frente a las empresas que aglutinan un mayor poder de mercado convierte la regulación europea en uno de los modelos con mayor intervención hacia el sector privado, en particular con exigencias que pueden limitar su potencial de crecimiento y que puede perjudicar a Europa en términos de mantener la carrera de la competitividad digital.

El segundo modelo es el protagonizado por los EE. UU. y que ha sido caracterizado por su menor regulación y la prominencia del sector privado y el *Big Tech*. Aunque ha habido varias iniciativas legislativas a nivel federal, como la Ley de Innovación Futura de IA o la Ley de Responsabilidad Algorítmica, ninguna ha prosperado por el momento. La regulación más importante es la Orden Ejecutiva presidencial 14110 de 2023 que establece algunos principios rectores para fomentar una mayor seguridad, fomento a la innovación y un modelo respetuoso con la privacidad en la IA, tanto a nivel nacional como a nivel internacional. sobre un Desarrollo. Aunque EE.UU. también presenta avances en regulación estatal, como la Ley de Transparencia de IA de California de 2024 que requiere a los sistemas de IA generativa que especifican cuándo el contenido es creado mediante esta técnica.

China presenta un modelo de mayor intervención bajo la preminencia del actor estatal. Una de los aspectos regulatorios más importantes son las Medidas Internas para la Administración de Servicios de IA, aprobadas en 2023. Frente a los otros modelos, en estas medidas se explicita que el contenido de los servicios de IA debe salvaguardar la seguridad nacional y proteger el interés público. Además de otros elementos en materia de privacidad y seguridad de los modelos de IA, se establece una revisión del contenido para evitar que este resulte inadecuado.

Como hemos mencionado, aunque no haya desarrollado un programa específico en materia de IA, el acercamiento del modelo de DPI constituye una alternativa con soluciones específicas para los países en desarrollo. Bajo el DPI, se privilegian sistemas de código abierto, bajo estándares que pueden ser compartidos entre los sectores privado y público para ir agregando capacidades a los sistemas digitales. Bajo el DPI se construyen también sistemas para la compartición de datos personales bajo estándares que pueden mejorar la transparencia en esta materia en el uso de IA.

VII. A MODO DE CONCLUSIÓN

La disrupción de la IA se da sobre un conjunto de dimensiones socioeconómicas demasiado amplio como para obviar una planificación regulatoria y de intervención estatal. Este capítulo se ha centrado sobre algunas de las dimensiones económicas afectadas, en particular atención a cuatro aspectos que han sido más claramente afectados por la IA y la economía digital en estas primeras etapas de su desarrollo. Los más férreos defensores no ya de una ausencia de intervención estatal, sino incluso de la eliminación de regulaciones

precautorias, reclaman que la complejidad intrínseca de la IA y su coyuntura de rápida aceleración hacen que los intentos regulatorios sean torpes y solo sean barreras a las potencialidades del desarrollo tecnológico.

Ante estos cambios estructurales de la economía, presentes y futuros, es preciso una mejor definición de la agencia en la dirección del cambio tecnológico, particularmente en los países del Sur Global, que ven rezagados sus legítimos intereses de alcance tecnológico frente a las agendas de los países líderes en materia de IA. Las dimensiones económicas afectadas por un cambio estructural como el de la IA pueden no ser intencionadas, derivado de una falta de control, pero la capacidad de contención depende del control de los Estados sobre la tecnología. El ejercicio de dicho control no es sencillo y no necesariamente el control debe entenderse como una barrera al desarrollo tecnológico.

Frente a las voces que reclaman regulaciones precautorias y una acción del Estado previsora de los profundos cambios socioeconómicos que derivan de la Revolución Digital, las *Big Tech* han favorecido la ausencia de regulación bajo todo tipo de argumentos entre los que destacan la enorme complejidad técnica de las tecnologías digitales y la indeterminación de posibilidades del desarrollo tecnológico, frente a la que la intervención estatal no puede intervenir eficazmente. Ciertamente, el abanico de posibilidades de la IA es difícil de abarcar en toda su potencialidad. Esto no debiera justificar la falta de acción, sino que es necesario un mayor estudio y más continuado de las implicaciones de la IA sobre la sociedad y, en particular, sobre la economía.

La IA tiene el potencial de avanzar exponencialmente la investigación científica, permitiéndonos de esta manera erradicar enfermedades, reducir los accidentes de tráfico a través del manejo autónomo, entre una lista interminable de posibles efectos beneficiosos. Sin embargo, en el otro extremo, entre el 37.8% y el 54.1% de investigadores especializados en IA otorgaban al menos un 10% de probabilidad de que los sistemas de IA desembocaran en eventos catastróficos, como la extinción humana (Grace *et al*. 2024). Nótese que estamos hablando de que más de un tercio de los expertos en el área consideran que existe una probabilidad nada desdeñable de terminar con los humanos como especie. Con respecto a impactos de carácter intermedio, temas como el desempleo tecnológico o la concentración empresarial son plenamente factibles, ni siquiera en el corto plazo, sino en el presente. Frente a estas posibilidades, que se materializan día a día, la política económica debe considerar sus respuestas frente a los retos de la IA, como un reto existencial. Cualquier otra actitud es una ilusión que no nos podemos permitir, menos desde los países en desarrollo.

VIII. BIBLIOGRAFÍA

Acemoglu, Daron y Pascual Restrepo. 2022. Tasks, automation and the rise in US wage inequality. *Econometrica* 90(5): 1973-2016.

Acemoglu, Daron y Pascual Restrepo. 2018. Artificial intelligence, automation and work. Working Paper 24196. *NBER working paper series*.

Crawford, Kate. 2021. *Atlas of AI*. New Haven y Londres: Yale University Press.

Diario Oficial de la Unión Europea. 2022a. Reglamento (UE) 2022/2065 de Servicios Digitales. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32022R2065

Diario Oficial de la Unión Europea. 2022b. Reglamento (UE) 2022/1925 de Mercados Digitales. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32022R1925

Eloundou, Tyna, Sam Manning, Pamela Mishkin, y Daniel Rock. 2023. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. arXiv:2303.10130

Freitag, Charlotte, Mike Berners-Lee, Kelly Widdicks, Bran Knowles, Gordon S. Blair y Adrian Friday. 2021. The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. *Patterns* 2(9) https://doi.org/10.1016/j.patter.2021.100340

Goodfellow, Ian J. *et al*. 2014. Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 27: 2672-2680.

Grace, Katja *et al*. 2024. Thousands of AI Authors on the Future of AI. AI impacts. https://aiimpacts.org/wp-content/uploads/2023/04/Thousands_of_AI_authors_on_the_future_of_AI.pdf

Human-Centered Artificial Intelligence. 2024. *Artificial Intelligence Index Report 2024*. Standford University.

Kangas, Olli, Signe Jauhiainen, Miska Simanainen, Minna Ylikännö.2021. *Experimenting with Unconditional Basic Income. Lessons from the Finnish BI Experiment 2017-2018*. Cheltenham (UK) y Northampton (MA, EE. UU.): Edward Elgar Publishing.

Krizhevsky, Alex, Ilya Sutskever. y Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25

Kurzweil, Ray. 2024. *The Singularity is Nearer. When We Merge with AI*. Nueva York: Viking.

Li, Pengfei, Jianyi Yang, Mohammad A. Islam y Shaolei Ren. 2023. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Fooprint of AI Models. https://arxiv.org/pdf/2304.03271

Luccioni, Alexandra Sasha, Sylvain Viguier y Anne-Laure Ligozat. 2022. Estimating the Carbon Footprint of BLOOM a 176B Parameter Language Model. https://arxiv.org/pdf/2211.02001

Parlamento Europeo. 2024. Artificial Intelligence Act. Consolidated text. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html#title2

Schmidhuber, Jürgen y Sepp Hochreiter. 1997. Long Short-Term Memory. *Neural Computation* 9(8): 1735-1780.

Tirole, Jean. 2024. Competition and industrial policy in the 21st century. *Oxford Open Economics* 3: i983-i1001.

Vanderbauwhede, Wim. 2023. Emissions from ChatGPT are much higher than from conventional search. https://limited.systems/articles/google-search-vs-chatgpt-emissions/

Vaswani, Ashish. *et al.* (2017) Attention Is All You Need. arXiv:1706.03762

Waymo. 2024.. Waymo Driver. Consultado en https://waymo.com/intl/es/waymo-driver/ el 25 de agosto de 2024.

CAPÍTULO CUATRO

INTELIGENCIA ARTIFICIAL. SU RELEVANCIA EN EL NUEVO DESARROLLO INDUSTRIAL DIGITAL

ESTRATEGIA PARA LA ADOPCIÓN DE LA IA EN LA INDUSTRIA MEXICANA

SUMARIO

I. Introducción **II.** La IA en la industria **III.** México ante la IA industrial **IV.** México en el Índice latinoamericano de Inteligencia Artificial (ILIA) **V.** Oportunidades de la IA en la industria **VI.** Gobernanza **VII.** Conclusiones **VIII.** Bibliografía

> JOSÉ LUIS SOLLEIRO
Investigador Titular del Instituto de Ciencias Aplicadas y Tecnologia (ICAT) de la UNAM

> ROSARIO CASTAÑÓN
Académica del Instituto de Ciencias Aplicadas y Tecnologías (ICAT) de la UNAM

> DAVID GUILLÉN
Gerente de consultoría y gestión de proyectos en Cambiotec, A.C.

I. INTRODUCCIÓN: LA EVOLUCIÓN DE LA INTELIGENCIA ARTIFICIAL

La inteligencia artificial (IA) ha evolucionado rápidamente desde sus inicios en la década de 1950, transformándose en una tecnología clave que impacta el desarrollo industrial y diversos aspectos de la vida cotidiana.

En 1936 Alan Turing, un matemático y criptógrafo británico, introdujo el concepto de la máquina de Turing, una idea teórica que formaliza los conceptos de algoritmo y computación, la cual fue fundamental para entender cómo funcionan las computadoras modernas[1]. En 1950, Turing propuso la famosa "Prueba de Turing" en su artículo "*Computing Machinery and Intelligence*". Esta prueba evalúa si una máquina puede exhibir un comportamiento inteligente indistinguible del de un ser humano. La propuesta de Turing sentó las bases para el desarrollo de la IA al plantear preguntas sobre la capacidad de las máquinas para pensar y razonar.

El término "inteligencia artificial" fue acuñado por John McCarthy en 1956 durante la Conferencia de Dartmouth, donde se sentaron las bases para el desarrollo de esta disciplina. Fue en esa época cuando se desarrollaron los primeros algoritmos y lenguajes de programación, como LISP[2], que permitieron a las máquinas realizar tareas simples basadas en reglas.

1 Durante la Segunda Guerra Mundial, Turing jugó un papel crucial en el desciframiento de los códigos nazis, especialmente los generados por la máquina Enigma

2 LISP, acrónimo de *list processing* (procesamiento de listas), es un lenguaje de programación que fue diseñado para una fácil manipulación de cadenas de datos. Es uno de los lenguajes de programación más antiguos todavía en uso.

En la década de los 60, la IA comenzó a generar algunas aplicaciones con el desarrollo de programas como ELIZA, un *chatbot* que simulaba conversaciones humanas, así como sistemas expertos como MYCIN[3], diseñado para diagnosticar enfermedades, lo que demostró el potencial de la IA en aplicaciones prácticas[4].

Los sistemas expertos (software que emula el razonamiento y la toma de decisiones de un experto humano en un área específica) dieron impulso a la IA. Uno de los primeros fue Dendral, creado en 1965 por Edward Feigenbaum y otros investigadores de la Universidad de Stanford, diseñado para el análisis de compuestos químicos[5]. Estos sistemas utilizan una base de conocimientos especializada y técnicas de razonamiento para analizar información y brindar soluciones a problemas específicos, imitando la toma de decisiones de un experto humano. En la década de los 80, los sistemas expertos evolucionaron y se aplicaron en diversos campos de la actividad económica.

Posteriormente, en la década de los 90, con el crecimiento de la capacidad de cómputo, se logró procesar algoritmos más complejos y simular decisiones. En un hito histórico, Deep Blue, una supercomputadora desarrollada por IBM, derrotó al campeón mundial de ajedrez Garry Kasparov en una partida oficial. Esto probó las capacidades avanzadas de cálculo y estrategia que podía conferírsele a las máquinas, abriendo paso a la identificación del potencial de la IA en la toma de decisiones[6].

A principios de este siglo, la llegada del *big data* (la capacidad de procesar grandes volúmenes de datos) revolucionó el campo de la IA. Los algoritmos de aprendizaje automático (*machine learning*) empezaron a evolucionar y consolidarse, permitiendo a las máquinas aprender patrones y hacer predicciones basadas en datos históricos. Este avance ha sido fundamental para aplicaciones como el reconocimiento facial y los motores de búsqueda.

Sobre esta base, el desarrollo del aprendizaje profundo (*deep learning*) llevó a la IA a nuevas aplicaciones más complejas. Utilizando redes neuronales

3 https://telefonicatech.com/blog/mycin-el-comienzo-de-la-inteligencia#:~:text=MYCIN%20era%20capaz%20de%20identificar,la%20meningitis%20y%20la%20bacteriemia consultada el 20 de noviembre de 2024

4 MYCIN era capaz de identificar las bacterias que causaban la infección en los pacientes y sugería los antibióticos y las dosis adecuadas para el peso de cada paciente. También era capaz de detectar enfermedades infecciosas de la sangre como la meningitis y la bacteriemia.

5 https://www.lifeder.com/sistemas-expertos/ consultada el 18 de noviembre de 2024

6 https://academiaready.es/los-hitos-mas-importantes-en-la-historia-de-la-inteligencia-artificial/ consultada el 18 de noviembre de 2024

artificiales con múltiples capas, esta técnica ha permitido avances significativos en áreas como el procesamiento del lenguaje natural (PLN), el reconocimiento de imágenes y los sistemas autónomos. Aplicaciones como Siri de Apple y el asistente virtual Alexa son ejemplos claros del impacto del aprendizaje profundo.

En 2012, la introducción de AlexNet, una red neuronal profunda que ganó el concurso ImageNet, marcó un avance significativo en el reconocimiento visual. Este modelo demostró que las redes neuronales[7] podían superar a los métodos tradicionales en tareas complejas como la clasificación de imágenes.

En 2020, DeepMind presentó AlphaFold, un sistema capaz de predecir con alta precisión la estructura tridimensional de proteínas a partir de su secuencia de aminoácidos. Este avance tiene implicaciones significativas para la biología molecular y el desarrollo de medicamentos.

Por otro lado, la inteligencia artificial generativa ha ganado popularidad recientemente, con herramientas como ChatGPT que pueden crear texto coherente y relevante a partir de indicaciones simples. Este tipo de IA está siendo utilizado en diversas industrias, desde marketing hasta desarrollo creativo. El crecimiento exponencial de la IA generativa, ha despertado preocupaciones sobre su uso ético y diversos riesgos, lo cual ha revelado la necesidad urgente de establecer regulaciones que garanticen la transparencia y equidad en los sistemas de IA para que se asegure su desarrollo responsable.

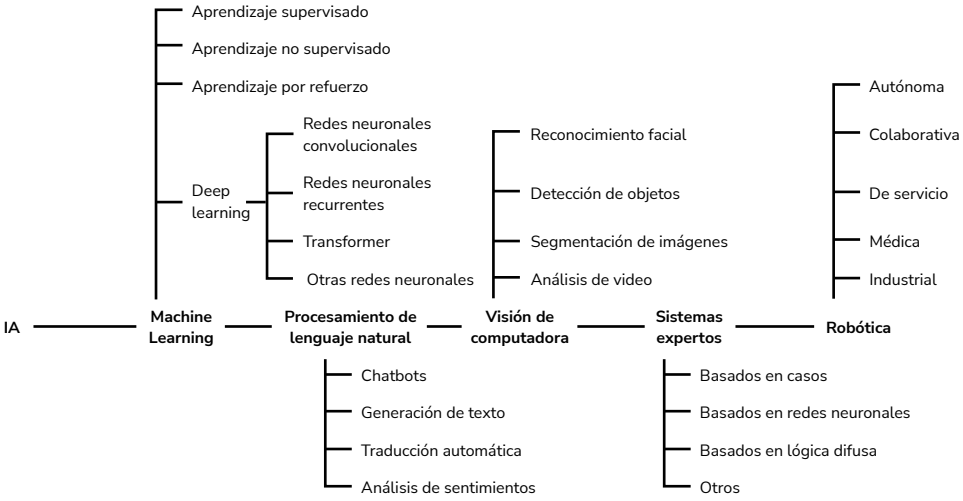
La IA abarca una amplia variedad de tecnologías, lo que hace que su taxonomía sea compleja y variada. Históricamente, se ha clasificado según diversos criterios, entre ellos:

- **Nivel de complejidad:** IA reactiva y limitada.
- **Capacidad para resolver problemas o nivel de inteligencia:** IA estrecha, general y superinteligencia.
- **Racionalidad y forma de operar:** sistemas que actúan como humanos, piensan como humanos o actúan de manera racional.

Además de estas clasificaciones, en la Figura 1 se muestran las principales disciplinas relacionadas con la IA.

7 Una red neuronal es un modelo computacional inspirado en el funcionamiento del cerebro humano. Está compuesta por unidades interconectadas llamadas neuronas artificiales que procesan información de manera similar a como lo hacen las neuronas biológicas.

FIGURA 1 Principales disciplinas relacionadas con la inteligencia artificial



FUENTE: Elaboración propia

En 2024, la inteligencia artificial continúa su expansión, siendo las empresas multinacionales las que más han aprovechado su potencial (Oldemeyer, 2024), (Sections, 2021). De acuerdo con la encuesta de (Szedlak, 2021), cerca del 90% de las PYMES aún no cuentan con aplicaciones de IA en sus operaciones. Debido a lo reciente de diferentes aplicaciones, aún existe una brecha significativa entre la aplicación de la tecnología para el desarrollo de soluciones concretas y su implementación en las empresas.

II. LA IA EN LA INDUSTRIA

Actualmente, términos como *lean manufacturing* (manufactura esbelta), manufactura avanzada e Industria 4.0 son bien conocidos en el ambiente industrial. De acuerdo con Lee (2020), todas estas técnicas han sido diseñadas para resolver eficientemente problemas conocidos, a ellas se ha sumado la IA, la cual utiliza diversos métodos de aprendizaje automático para integrar diferentes tipos de datos permitiendo extraer datos invisibles, identificar relaciones y optimizar sistemas, evitando así problemas (ocultos o desconocidos) que aún no han ocurrido.

La IA industrial es un nuevo campo emergente que se está expandiendo a un ritmo muy rápido, gracias a la confluencia de múltiples sensores integrados en diversos productos, la creciente disponibilidad de datos (particularmente datos

digitalizados), algoritmos más inteligentes, y la disponibilidad de procesadores y dispositivos de almacenamiento de grandes volúmenes de datos.

De acuerdo con lo anterior, es importante tener en mente que la IA es una ciencia cognitiva que incluye principalmente seis campos: procesamiento de lenguaje natural; visión por computadora; cognición y razonamiento; teoría y ética de juegos; y aprendizaje automático y robótica.

Los expertos tradicionales en IA han desplegado su gran capacidad algorítmica, pero esto no es suficiente para las aplicaciones industriales, pues es difícil resolver problemas de sistemas industriales en ausencia de expertos y conocimiento del entorno empresarial en diversos sectores. Por ello, el mayor desafío de la IA industrial es transformar el pensamiento algorítmico centrado en el ser humano en la comprensión de sistemas de ingeniería.

El desarrollo de sistemas de fabricación inteligente y la evolución de Internet han permitido ver una tendencia en la que los conceptos de "inteligencia localizada" e "inteligencia interconectada" van de la mano en el contexto de los sistemas industriales los cuales están evolucionando hacia las aplicaciones inteligentes. En este sentido, uno de los mayores desafíos para la IA industrial no radica en la tecnología en si misma, sino en cómo aplicarla para generar valor. Se trata entonces de usar la IA para resolver problemas como los siguientes:

- Calidad en la manufactura, incluyendo equipo, sistemas y procesos, así como la organización industrial y la construcción de capacidades del personal. Con modelos de la IA, se puede contar con una detección temprana y automatizada de posibles defectos durante la producción utilizando datos en tiempo real, por ejemplo en ensamblaje y mecanizado. Las aplicaciones emergentes incluyen la inspección visual automatizada utilizando métodos de aprendizaje profundo (Ojer, 2020), la predicción de defectos (en línea con el paradigma de fabricación de cero defectos) (Peres, 2019) y la predicción de calidad en línea (Schmitt, 2018)
- Estandarización de los flujos de trabajo, con lo que se busca optimizar y uniformizar las tareas y procesos dentro de una organización, aprovechando las capacidades de la IA para automatizar y mejorar la eficiencia.
- Mantenimiento predictivo utilizando algoritmos para analizar datos en tiempo real provenientes de sensores instalados en maquinaria y equipos, lo cual permite predecir fallos antes de que ocurran y programar mantenimientos oportunos.
- Optimización de cadenas de suministro mediante el análisis avanzado de datos, para ajustar niveles de inventario y mejorar la logística.

En palabras de Lee (2020, p. 9), el objetivo del futuro sistema de Inteligencia Industrial es "crear un entorno industrial que alcance las metas de cero accidentes, cero contaminación, cero residuos, cero defectos y cero tiempo de inactividad".

CAPACIDADES REQUERIDAS PARA LA IA INDUSTRIAL.

El desarrollo de la IA industrial requiere una serie de capacidades tecnológicas que hagan viable su implementación efectiva y su integración en los procesos existentes. A continuación, se detallan las principales:

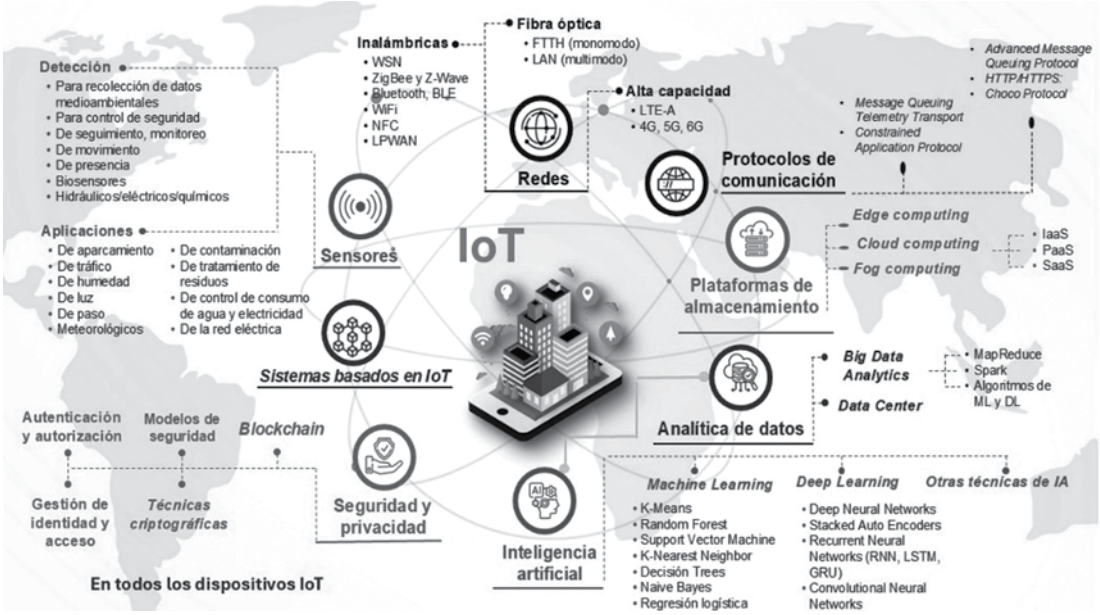
- Infraestructura de datos, la cual implica sistemas que permitan la recolección continua de datos a partir de sensores, máquinas y procesos. Esto requiere bases de datos robustas, precisas y actualizadas, así como la capacidad de almacenamiento de grandes volúmenes de información.
- Herramientas avanzadas de análisis de datos. Se requiere software de análisis estadístico y plataformas de *big data* que permitan procesar y analizar grandes conjuntos de datos en tiempo real. De igual forma, se necesita capacidad para desarrollar y aplicar algoritmos de aprendizaje automático (machine learning) que permitan a las máquinas aprender y adaptarse a partir de los datos para tomar decisiones de forma autónoma.
- Capacidades de automatización mediante sistemas para controlar y supervisar procesos, como sistemas SCADA (Control de Supervisión y Adquisición de Datos), LiDAR (Light Detection and Ranging) y PLC (Controladores Lógicos Programables).
- Robótica avanzada mediante robots colaborativos capaces de trabajar con humanos y también de manera autónoma. Actualmente, la integración de la inteligencia artificial (IA) y las tecnologías del Internet de las Cosas (IoT) está llevando a los robots industriales a nuevos niveles de autonomía y adaptabilidad. Estos avances permiten a los robots aprender de su entorno y optimizar sus operaciones, lo que resulta en una mayor eficiencia en los procesos de manufactura (Schwab, 2017).
- Internet de las Cosas (IoT). La Internet de las cosas (IoT) describe la red de objetos físicos ("cosas") que llevan incorporados sensores, software y otras tecnologías con el fin de conectarse e intercambiar datos con otros dispositivos y sistemas a través de Internet. Estos dispositivos van desde objetos domésticos comunes hasta herramientas industriales sofisticadas[8].

8 https://www.oracle.com/mx/internet-of-things/what-is-iot/ consultada el 15 de agosto de 2023

La conectividad entre dispositivos es fundamental para la recopilación y el intercambio de datos para generar. En la Figura 2 se aprecian los componentes de infraestructura para el funcionamiento de la IoT.

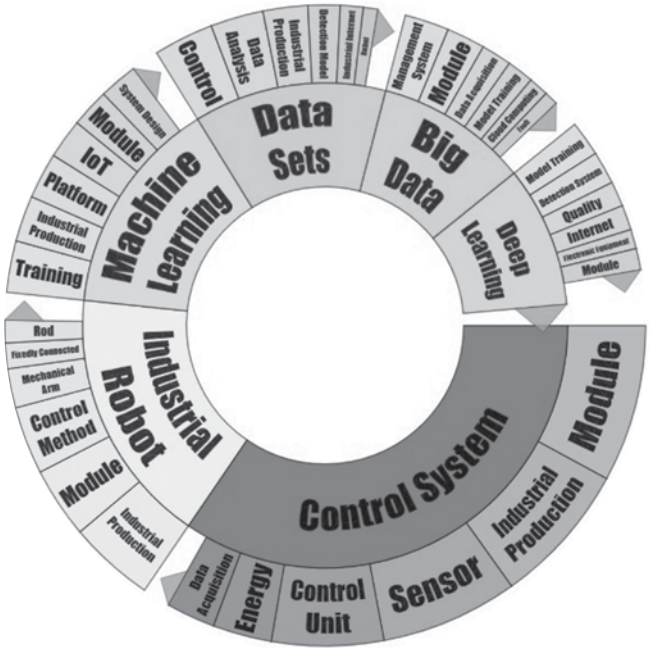
- Redes 5G para mejorar la capacidad para transmitir grandes volúmenes de datos rápidamente y con baja latencia, lo cual es crucial para aplicaciones industriales que requieren respuesta inmediata, como son los sistemas automatizados y los robots industriales.
- Ciberseguridad y protección de datos para implementar regulaciones y medidas robustas para proteger los datos industriales y las infraestructuras críticas. Sistemas efectivos para identificar amenazas y reaccionar ante ataques cibernéticos son esenciales para lograr la operación confiable en un contexto de alta conectividad en el que cada punto de contacto en una red es un elemento de riesgo (Solleiro et al., 2022).
- Capacidades de investigación y desarrollo en áreas como redes neuronales, aprendizaje profundo, aprendizaje de refuerzo, métodos estadísticos, control inteligente, sistemas ciberfísicos, sensores, simulación y ciberseguridad. En la Figura 3, se muestran las principales tecnologías protegidas por patentes en el área de IA en el periodo 2014-2024. Es claro que los grandes campos donde se concentran las invenciones son: aprendizaje automático, análisis de datos, computación en la nube, aprendizaje profundo, sistemas de control y robots industriales.

FIGURA 2 Los elementos de infraestructura para el funcionamiento de IoT



FUENTE: Elaboración propia

FIGURA 3 Áreas de concentración de las invenciones en IA industrial (periodo 2014-2024)



FUENTE: Elaboración propia utilizando Patbase

Como puede observarse, la aplicación efectiva de la IA industrial depende del desarrollo e integración de capacidades tecnológicas que incluyen desde una infraestructura sólida para la recolección, análisis y uso de datos hasta la automatización avanzada, la ciberseguridad y la investigación. Sólo con base en esas capacidades, las empresas lograrán la asimilación plena de estas tecnologías. Las inversiones requeridas para construir estas capacidades no se concentran sólo en herramientas tecnológicas, sino también en la capacitación del personal y en la creación de una cultura organizacional que valore la innovación y el aprendizaje continuo.

III. MÉXICO ANTE LA IA INDUSTRIAL

En el contexto mexicano, la adopción de la IA no sólo presenta oportunidades para mejorar la productividad y la competitividad de sus empresas, sino que también una serie de obstáculos estructurales derivados de un sistema nacional de innovación poco articulado y políticas públicas inadecuadas que no han colocado el desarrollo de la economía digital como uno de sus ejes.

De acuerdo con el Índice de Preparación para la IA del Fondo Monetario Internacional (FMI), México ocupa el puesto 56 de 174 países, con un puntaje de 0.53. Este índice evalúa varios componentes críticos, incluyendo infraestructura digital, capital humano, innovación e integración económica, y regulación y ética. En comparación con otros países de América Latina, México se sitúa detrás de Chile (0.58) y Uruguay (0.55) en términos de preparación para adoptar esta tecnología[9].

INFRAESTRUCTURA

En la dimensión de infraestructura digital, México obtuvo un puntaje de 0.13, lo que refleja deficiencias significativas en acceso a internet, velocidad de banda ancha y cobertura móvil. La infraestructura inadecuada es una limitante seria para el desarrollo de la IA industrial, pues las aplicaciones en este campo requieren alta velocidad y baja latencia, las cuales dependen de una red 5G extendida.

De acuerdo con Piedras (2024), en la primera mitad de 2024, "el balance del presupuesto público asignado y ejecutado en el sector TIC en México revela un panorama contrastante: algunas instituciones o programas recibieron un impulso significativo, mientras que otros enfrentan recortes que podrían limitar su alcance". La Comisión Federal de Electricidad y PEMEX tuvieron aumentos notables: CFE 339.4% en servicios de infraestructura de telecomunicaciones y 203.5% para los servicios de telecomunicaciones de PEMEX. En franco contraste, la Secretaría de Infraestructura, Comunicaciones y Transportes tuvo una reducción del 19.8% en ese periodo, y la Dirección General @prende.mx (anteriormente Dirección General de Educación Televisiva), programa de la Secretaría de Educación Pública (SEP) enfocado en la digitalización educativa y la conectividad en escuelas, tuvo una reducción de 15.1%, lo cual deja ver que el discurso de gubernamental no coincide con su compromiso relativo a la educación digital. Lo destacable de estos datos presupuestales es que revelan cuáles han sido las prioridades establecidas.

En el caso del despliegue de 5G, México ha aumentado su cobertura en zonas urbanas, no así fuera de ellas[10]. Pero el déficit de infraestructura 5G no es su único desafío, pues el alto costo del espectro ha entorpecido el proceso de su asignación, lo que retrasa aún más el desarrollo de esta red.

9 https://www.eleconomista.com.mx/opinion/El-futuro-digital-de-Mexico-Oportunidades-y-retos-en-IA-20240718-0003.html

10 https://www.5gamericas.org/resources/deployments/ consultada el 20 de noviembre de 2024

CAPITAL HUMANO PARA IA

La inversión en educación, especialmente en áreas STEM (ciencia, tecnología, ingeniería y matemáticas), es insuficiente. Se requiere una mayor capacitación y reentrenamiento para la fuerza laboral actual. Un estudio del Instituto Mexicano para la Competitividad (IMCO, 2023) señala que la falta de personal capacitado es uno de los principales obstáculos para adoptar plenamente la IA.

A pesar de la creciente demanda, la oferta de profesionales especializados en áreas como aprendizaje automático, procesamiento de datos y desarrollo de algoritmos sigue siendo limitada. Este déficit pone en riesgo la capacidad del país para aprovechar las ventajas de la inteligencia artificial y lo expone a una mayor dependencia tecnológica de mercados extranjeros.

CISCO (2024), con base en una encuesta global, concluye que en 2024, sólo el 27% de las empresas en México considera tener personal capacitado para enfrentar los retos en inteligencia artificial, una cifra que representa una caída respecto al año anterior. Además, se detecta una creciente resistencia cultural a nivel organizacional, especialmente en niveles gerenciales y operativos.

IA EN LAS EMPRESAS

El reciente informe de Cisco sobre el *Índice de Preparación en Inteligencia Artificial 2024*[11] revela que la capacidad de las empresas mexicanas para implementar IA se ha estancado. El estudio, que analiza 8,000 compañías en 30 países, clasifica a las organizaciones en cuatro niveles de preparación: líderes, cazadores, seguidores y rezagados. En el caso de México, solo el 13% de las empresas se encuentran en la categoría de líderes, mientras que un 38% están en el nivel de cazadores, ligeramente por encima del promedio global.

En comparación con 2023, se observa una disminución en el porcentaje de empresas que se consideran líderes en el uso de IA. Mientras que el número de cazadores se mantiene estable.Las organizaciones clasificadas en los niveles más bajos, seguidores y rezagados, han aumentado, hecho que evidencia un retroceso en la preparación nacional para enfrentar los desafíos de la inteligencia artificial.

"Según los datos, uno de los mayores obstáculos en México es la infraestructura inadecuada, lo que incluye deficiencias en cómputo, redes y ciberseguridad. Solo el 21% de las empresas cuenta con los recursos tecnológicos

11 https://comentariodeldia.com/2024/11/25/mexico-retrocede-en-preparacion-para-la-ia-segun-informe-de-cisco/ consultada el 20 de noviembre de 2024

necesarios, un descenso considerable respecto al 28% registrado en 2023. Además, el problema de la gestión de datos sigue siendo significativo, con el 82% de las compañías trabajando con información fragmentada en silos, lo que limita la calidad y fiabilidad de sus modelos de IA"[12].

INNOVACIÓN

En esta dimensión, México ha retrocedido. El gasto en investigación< es bajo en comparación con estándares internacionales, alcanzando solo el 0.27% del PIB[13] en 2023, cuando hace seis años rondaba el 0.5%. Los Programas Nacionales Estratégicos definidos por CONAHCYT en la administración de Andrés Manuel López Obrador no consideraron prioritaria la investigación para la industria manufacturera, además de que los fondos sectoriales y los mixtos fueron desaparecidos. En lo relativo al fomento del desarrollo tecnológico en las empresas, también fueron eliminados los apoyos del Programa de Estímulos a la Innovación y los que manejaba el Instituto Nacional del Emprendedor.

En los Encuentros IA Hoja de ruta para México (May y Rojas, 2024, p. 28), se menciona, para el rubro de investigación y desarrollo, que "el grupo de participantes destacó que la inversión de México en I+D se encuentra rezagada en comparación con otros países de Latinoamérica, lo cual se refleja en un ecosistema de investigación disminuido, con menos investigadores e investigadoras, menor productividad y una menor cantidad de artículos y patentes registrados". También May y Rojas (2024, p. 29) subrayan que "los desafíos específicos que enfrentan las PyMEs en el ámbito de la I+D son enormes dado que con frecuencia carecen de los recursos financieros y la capacidad de investigación interna necesarios. El grupo de personas expertas reconoció la ausencia, hasta el momento, de cualquier iniciativa, fondo o programa destinado a la investigación y desarrollo tecnológico dirigido específicamente a estas empresas o la colaboración multisectorial. Este aspecto fue subrayado como fundamental, dado que las PyMEs constituyen más del 95% de la economía del país".

En estas circunstancias, la capacidad de innovación del país y sus empresas no sólo es baja sino que también ha sido mermada.

12 https://comentariodeldia.com/2024/11/25/mexico-retrocede-en-preparacion-para-la-ia-segun-informe-de-cisco/

13 https://www.elfinanciero.com.mx/opinion/eugenio-gomez/2024/08/27/que-tan-bien-preparado-esta-mexico-para-adoptar-la-inteligencia-artificial/

REGULACIÓN Y ÉTICA.

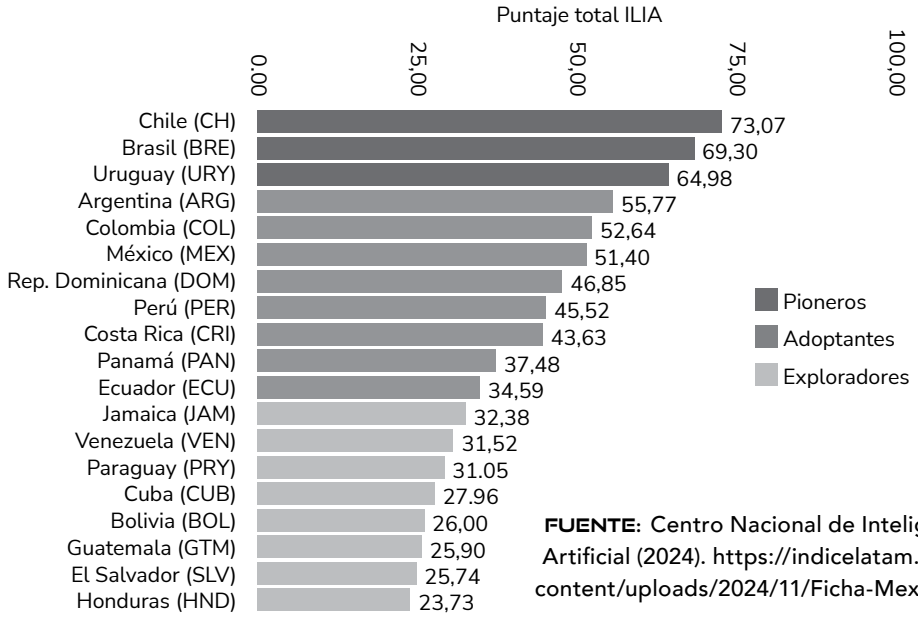
La falta de un marco regulatorio adaptado a las necesidades del entorno digital presenta un obstáculo adicional para la adopción efectiva de la IA. México requiere desarrollar un marco regulatorio flexible que proteja los derechos ciudadanos mientras fomenta la innovación. Lo anterior puede facilitar una adopción más rápida y efectiva de tecnologías basadas en IA. May y Rojas (2024) reconocen que México requiere construir una Estrategia Nacional de IA basada en principios éticos de la tecnología, así como designar una entidad responsable para la implementación de la Estrategia Nacional de IA.

IV. MÉXICO EN EL ÍNDICE LATINOAMERICANO DE INTELIGENCIA ARTIFICIAL (ILIA)

En 2024, se lanzó la segunda edición del ILIA con el fin de ofrecer una visión integral y estructurada sobre el progreso de la IA en América Latina y el Caribe, con base en elementos y variables que inciden en el desarrollo de los ecosistemas de IA en torno a tres dimensiones: factores habilitantes, investigación, desarrollo y adopción, y gobernanza.

La Figura 4 ilustra que, aún en el contexto de esta región, México se ha rezagado ocupando apenas el sexto lugar en una clasificación de "adoptante" de IA, por su grado de madurez.

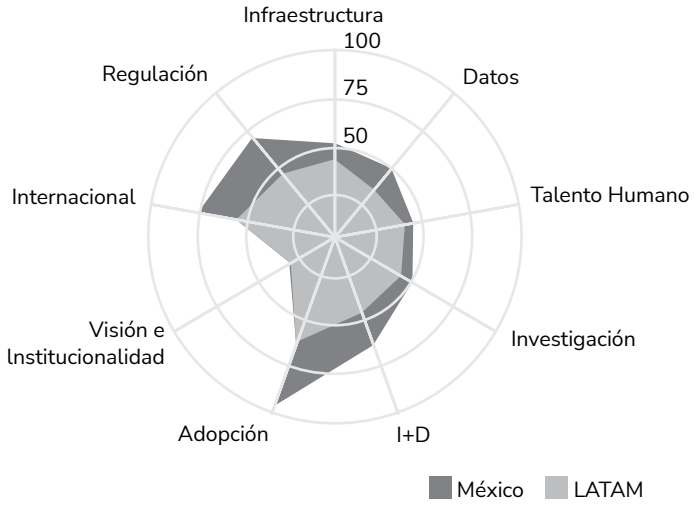
FIGURA 4 El Índice Latinoamericano de IA en 2024



FUENTE: Centro Nacional de Inteligencia Artificial (2024). https://indicelatam.cl/wp-content/uploads/2024/11/Ficha-Mexico.pdf

En la Figura 5, se observa la posición de México frente al promedio de los países de América Latina y el Caribe, revelando un desempeño superior, aunque con un perfil similar. No obstante, resalta la falta de visión e institucionalidad de México en el ámbito de la IA, lo que lo ubica en franca desventaja.

FIGURA 5 Subdimensiones de México y LATAM en el ILIA 2024



FUENTE: Centro Nacional de Inteligencia Artificial (2024). https://indicelatam.cl/wp-content/uploads/2024/11/Ficha-Mexico.pdf

En el documento "Hacia una estrategia nacional de inteligencia artificial" (Ibarra et al., 2024, p. 48), se presenta una revisión general de las Estrategias Digitales Nacionales recientes, concluyendo que "la falta de visión a largo plazo de las políticas públicas y falta de continuidad de políticas alineadas a buenas prácticas o estándares internacionales en materia de digitalización, son un grave problema y un obstáculo para que México alcance crecimientos significativos en materia de IA. Esta situación se repite cada seis años en la administración pública federal y locales, y cada 3 años en las administraciones municipales o de alcaldías de la Ciudad de México".

La Alianza Nacional de Inteligencia Artificial (Lagunes et al., 2024, p. 12), menciona que, en México, "la adopción de las tecnologías basadas en la IA muestra una tendencia al alza. El "Global AI Adoption Index 2022" señala que 31% de las empresas en México han implementado activamente soluciones con IA (IBM Global AI Adoption Index 2022, 2022). Además, una encuesta realizada por la Asociación de

Internet y Knowsy AI (2023) indica que 55% de los usuarios de internet en México perciben la IA como una herramienta útil, mientras que 22% expresa preocupaciones sobre sus posibles impactos negativos. Estas cifras reflejan una dualidad presente tanto en México como a nivel global: un entusiasmo por la innovación y el desarrollo tecnológico, junto con una cautela hacia los riesgos éticos, económicos, sociales, de dependencia tecnológica y de ciberseguridad que estos avances podrían conllevar".

Ante esa dualidad, en México se han lanzado diversas propuestas para definir políticas de IA, las cuales comparten su orientación regulatoria, relegando el componente de fomento. También es notorio que recientemente se ha puesto énfasis en la IA generativa y su uso ético y responsable. Tal es el caso de la Alianza Nacional de Inteligencia Artificial (ANIA) gestada en el Senado de la República, cuya propuesta para una agenda de política se resume en la Figura 6. Es evidente que el impulso al desarrollo y aplicación de la IA industrial no tiene un eje específico.

FIGURA 6 Ejes de la propuesta de la ANIA para construir una agenda nacional de IA 2024-2030

FUENTE: Lagunes *et al*. (2024)

V. OPORTUNIDADES DE LA IA EN LA INDUSTRIA

Una política de fomento a la IA industrial debería orientarse a impulsar sus aplicaciones más prometedoras, las cuales incluyen varias áreas innovadoras que buscan mejorar la eficiencia, la toma de decisiones y las capacidades operativas. Se espera que entre las aplicaciones de mayor impacto para el corto y mediano plazo se encuentren las siguientes:

OPTIMIZACIÓN DE PROCESOS EN LA FABRICACIÓN

Con la aplicación de IA industrial para la optimización de procesos se busca aumentar la rentabilidad de procesos, al mismo tiempo que más sostenibles y eficientes.

Se trata de sistemas de apoyo a la toma de decisiones en tiempo real a partir de modelos basados en datos históricos. Ejemplos de soluciones tangibles que se están desarrollando a nivel mundial son:

- Predicción y optimización (gestión) del consumo de energía (Qin, 2018)
- Eficiencia de la producción (Liang, 2019)
- Previsión de la demanda (Chien, 2020).
- Optimización del diseño y producción de piezas complejas, la creación rápida de prototipos y la personalización (principalmente mediante impresión y fabricación 3D) (Liu, 2024)

CONTROL Y ASEGURAMIENTO DE CALIDAD.

Con modelos de IA, contar con una detección temprana y automatizada de posibles defectos durante la producción utilizando datos en tiempo real, por ejemplo en ensamblaje y mecanizado. Las aplicaciones emergentes incluyen la inspección visual automatizada utilizando métodos de aprendizaje profundo (Ojer, 2020), la predicción de defectos (en línea con el paradigma de fabricación de cero defectos) (Peres, 2019) y la predicción de calidad en línea (Schmitt, 2018).

COLABORACIÓN ENTRE HUMANOS Y ROBOTS

Actualmente, la integración de la IA y las tecnologías del IoT está llevando a los robots industriales a nuevos niveles de autonomía y adaptabilidad. Los cobots, o robots colaborativos, están diseñados específicamente para trabajar

de forma segura y conjunta con los humanos, lo que representa un cambio importante respecto a los robots industriales tradicionales, que operaban de manera aislada debido a los riesgos de seguridad. Los cobots están equipados con sensores avanzados que les permiten detectar la presencia de humanos y ajustar sus movimientos para evitar accidentes. Se brinda apoyo a las tareas actuales centradas en humanos, ya sea mejorando el bienestar y la seguridad de los operadores o haciendo que sus tareas sean más fáciles y eficientes. Las aplicaciones de la IA industrial incluyen la capacitación de la fuerza laboral y el apoyo a las tareas (Lai, 2020), la robótica colaborativa (Bergamini, 2020) y la ergonomía (Abobakr, 2019).

SISTEMAS EXPERTOS PARA LA TOMA DE DECISIONES

El uso de sistemas expertos en entornos industriales permite la toma de decisiones automatizadas basada en reglas y bases de conocimiento establecidas. Estos sistemas se aplican en diversas áreas como la evaluación de riesgos, el diagnóstico de fallas y el mantenimiento predictivo (Yan, 2017).

OPTIMIZACIÓN DE LA LOGÍSTICA Y LA CADENA DE SUMINISTRO

Mediante la automatización y el análisis predictivo se puede lograr la articulación entre los eslabones de una cadena de suministro, optimizando la comunicación, estandarización y manejo de datos críticos (Oran, 2021). El uso de IA ayuda a superar barreras geográficas y de lenguaje y a generar ambientes colaborativos muy eficientes.

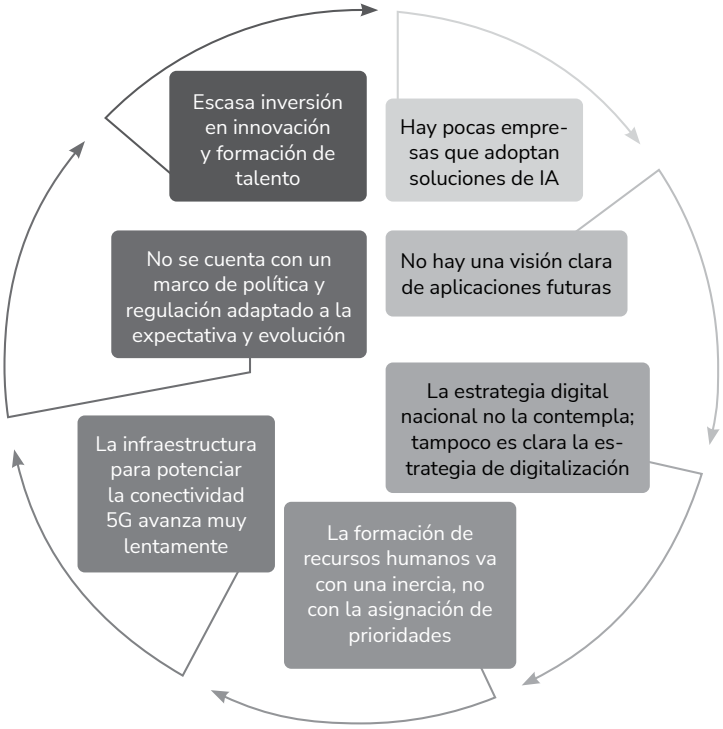
LA NECESIDAD DE UNA POLÍTICA DE IA INDUSTRIAL

Para ir al encuentro de las oportunidades expuestas, se requiere una política clara, con visión de largo plazo que ayude a superar las carencias actuales, sintetizadas en la Figura 7.

Para superar estos obstáculos, la política de IA industrial debe tomar en cuenta al menos los siguientes elementos:

Apoyo financiero a las empresas que desarrollen y adopten esta tecnología, mediante incentivos económicos directos en la modalidad de financiamiento parcial en función del mérito de la propuesta empresarial, así como estímulos fiscales a la inversión en este campo. Las PYME deberían recibir tratamiento

FIGURA 7 Obstáculos al desarrollo y adopción de la IA industrial en México



FUENTE: Elaboración propia

preferencial. Debe tenerse en cuenta que los costos de la tecnología y la habilitación del personal suelen ser altos, lo cual debe considerarse en el diseño del instrumento de apoyo.

Deben generarse fondos específicos para propiciar la investigación conjunta entre centros de investigación y las empresas. Estos fondos se dedicarían a financiar proyectos innovadores que ofrezcan soluciones prácticas para la implementación de soluciones de IA e internet de las cosas en empresas[14]. La integración de la IA con otras tecnologías emergentes como la computación cuántica y la robótica promete expandir aún más sus capacidades para resolver problemas complejos e impactar sectores económicos diversos.

14 En la propuesta de Agenda Nacional de IA para México 2024-2030, se sugiere focalizar fondos de inversión en desarrollo de casos de uso de IA en industrias cuyos equipos directivos son diversos en términos de experiencia profesional y paritarios en el número de mujeres en espacios de liderazgo. Los casos de uso podrán utilizar infraestructura pública digital y espacios controlados de prueba *sandboxes*, así como, acceso a redes de investigación y desarrollo nacionales e internacionales

Construir infraestructura de conectividad de alta velocidad y baja latencia accesible en las áreas de desarrollo industrial y facilitar el acceso a los servicios de internet a costos accesibles. Para lograr estas características, es fundamental que se trabaje en el marco de arreglos público-privados con visión de largo plazo.

De igual manera, debe haber un suministro confiable de energía eléctrica para poder operar de manera confiable la red de dispositivos inteligentes interconectados, así como los centros de datos.

Respecto a esto último, coincidimos con la propuesta de Lagunes *et al*. (2024) en relación con la infraestructura de datos, pues "se requiere promover la inversión en el desarrollo de Infraestructura Pública Digital (DPI), incluyendo redes de telecomunicaciones, software, modelos de IA, y estructuras de datos abiertos diseñadas bajo principios éticos". Para ello, es necesario definir un marco normativo y de fomento que facilite la creación y operación de centros de datos compartidos, considerando estándares de seguridad, ética y eficiencia energética. Operar un centro de datos compartido requiere una inversión significativa en infraestructura, tecnología, personal y procesos, pues es fundamental garantizar la seguridad, la disponibilidad y la calidad del servicio para satisfacer las demandas de los clientes y mantener la competitividad en el mercado. Este es un espacio adicional para la promoción de alianzas público-privadas.

Es necesario invertir en la creación y consolidación de plataformas de análisis de datos, con software para su procesamiento y análisis mediante herramientas avanzadas, para que las empresas puedan procesar grandes volúmenes de datos generados por dispositivos IoT, operaciones de comercialización y logística.

La ciberseguridad y la protección de la privacidad y los datos personales son elementos necesarios para la aplicación responsable de la IA. México enfrenta retos muy importantes en esta materia, pues es un destino frecuente para los ciberataques. No se cuenta con una estrategia nacional ni con una legislación específica a pesar de que las iniciativas han proliferado. En ese sentido, se recomienda considerar los planteamientos de la Guía para la elaboración de una estrategia de ciberseguridad nacional promovida por el Instituto Federal de Telecomunicaciones (Solleiro y Cuevas, 2024).

Incrementar las inversiones en educación técnica y profesional es indispensable para cerrar la brecha de habilidades necesarias para implementar IA en la industria. Fomentar programas educativos enfocados en STEM es crucial para preparar a las futuras generaciones. Se requiere, de acuerdo con Lagunes et al. (2024), un ecosistema que contemple los elementos clave que permita la integración de las tecnologías basadas en la IA dentro de los entornos educativos (infraestructura, desarrollo profesional docente, oferta de recursos

educativos digitales, así como mecanismos de monitoreo y evaluación). Este planteamiento de política implica la definición de una estrategia educativa de largo plazo que rompa con las inercias del actual sistema educativo.

VI. GOBERNANZA

La gobernanza se refiere al conjunto de procesos, estructuras y mecanismos que se utilizan para tomar decisiones de forma participativa e inclusiva para implementar políticas y gestionar recursos en el marco de la política. Uno de sus objetivos es garantizar que la estrategia de IA se integre en el diseño organizacional, considerando a los actores del ecosistema industrial. Este es un punto de partida para el diseño, desarrollo e implementación de una gobernanza cibernética.[15] La política de IA debe basarse en un marco de gobernanza participativa para fomentar el diálogo multisectorial. Para este efecto, es necesario construir mecanismos de análisis, consulta y decisión propios de los sistemas de gobernanza abierta y participativa. Los mecanismos de comunicación y mejora continua deben incluir el diálogo constante con organizaciones de disciplinas que tienen relación e incidencia en la aplicación de la IA. Es importante seguir los principios siguientes:

- Participación de los diversos grupos de interés en el proceso de toma de decisiones para definir, ejecutar y evaluar la estrategia nacional de IA, asegurando que se consideren diversas perspectivas y necesidades.
- Transparencia para que las decisiones, acciones, avances y resultados sean accesibles para todas las partes interesadas, promoviendo la confianza y la rendición de cuentas.
- Responsabilidad compartida y distribución de competencias para asegurar que las personas y entidades realicen las actividades que les corresponden y rindan cuentas por sus decisiones y comportamientos.
- Eficiencia y eficacia para que los recursos destinados a la realización de la Estrategia se utilicen de manera óptima para lograr los objetivos establecidos, minimizando el desperdicio y maximizando los resultados.
- Equidad para evitar la discriminación de grupos, organizaciones y tecnologías, promoviendo la inclusión, la neutralidad tecnológica y la igualdad de oportunidades.

15 En la presente administración el Instituto Federal de Telecomunicaciones dejó de ser un órgano autónomo, pasará a integrarse a la Secretaría de Economía. Aún cuando se ha asegurado que esta acción no afectarán las acciones que se estaban realizando, lo cierto es que tal afirmación no puede aún ser verificada.

La Presidenta de México, Claudia Sheinbaum Pardo, presentó en noviembre de 2024 la nueva Agencia de Transformación Digital y Telecomunicaciones, cuya principal misión será digitalizar los trámites que hoy se realizan de forma presencial, lo que permitirá continuar con una de las principales prioridades de la Cuarta Transformación: el combate a la corrupción. Para simplificar los trámites que tienen que hacer los ciudadanos, se creará Llave MX, un nuevo mecanismo de autentificación digital que permitirá agilizarlos. Como puede observarse, esta Agencia tiene una vocación hacia dentro del gobierno y, hasta el momento, no contempla asumir la tarea de generar una estrategia en materia de IA que pueda abordar las cuestiones asociadas a una política industrial digital. En este contexto, cobra relevancia la propuesta de Lagunes et al. (2024, p. 16) en el sentido de que un objetivo principal de la agenda para México es "incorporar acciones de política pública en Inteligencia Artificial como parte del Plan Nacional de Desarrollo (PND), la Estrategia Digital Nacional (EDN), y el Plan de Cumplimiento de las Metas de Desarrollo Sostenible 2030,definiendo una hoja de ruta por acción, instancias responsables e indicadores de ejecución". No es sostenible que México trate de incorporarse a esta ola tecnológica tan trascendente sin tener visión, estrategia y gobernanza sólida.

VII. CONCLUSIONES

Si bien México muestra un alto potencial para adoptar inteligencia artificial en la industria, principalmente por la operación de sectores manufactureros integrados a mercados de exportación y por las posibilidades del nearshoring, el país enfrenta desafíos significativos que deben ser abordados urgentemente. Sin duda, el principal es la definición de una estrategia de desarrollo y adopción amplia de la IA en las empresas, particularmente las pyme. La estrategia debe corresponder con una política pública acompañada de instrumentos efectivos.

Entre los instrumentos de la política de IA destacan la legislación y regulación efectivas. La ausencia de un marco regulatorio claro sobre el desarrollo y uso de la IA puede generar incertidumbre legal y operativa para las empresas. Esto es especialmente relevante en México donde no existen actualmente leyes específicas que regulen el uso ético y responsable de la inteligencia artificial, así como cuestiones relacionadas con la ciberseguridad y la privacidad. La desaparición de organismos autónomos como el INAI y el IFT genera incertidumbre, por lo que se debe actuar con premura para ofrecer claridad en estas materias.

La inversión en infraestructura digital, el refuerzo de la educación en sus distintos niveles para poder contar con talento humano en las materias críticas

y la atención a la investigación, desarrollo tecnológico e innovación serán elementos fundamentales para mejorar el desempeño de México en este campo.

Hay que tener claridad sobre la necesidad de inversión pública y privada para que los programas y proyectos para ejecutar la estrategia sean viables. El gobierno no podrá afrontar los desafíos solo, por lo que debe tener gran apertura a la colaboración con el sector privado, con una perspectiva de largo plazo. Esto requiere un ambiente de seguridad jurídica que actualmente no se está ofreciendo.

Con un enfoque estratégico y colaborativo, México puede no sólo reducir las brechas existentes en las clasificaciones internacionales, sino también lograr que las empresas del país tengan una incorporación virtuosa en la inteligencia artificial y alcancen un nivel de dominio tal que les permita extraer el máximo valor de las tecnologías asociadas.

La adopción efectiva de la IA en el sector industrial requiere un enfoque integral que combine tecnología avanzada, gestión eficiente de datos, capacitación del personal y una cultura organizacional abierta a la innovación. Al desarrollar estas capacidades, las empresas pueden no solo mejorar su eficiencia operativa, sino también posicionarse competitivamente en un mercado cada vez más digitalizado.

La implementación de proyectos de IA puede ser costosa, incluyendo gastos en hardware, software y personal técnico especializado. Esto puede ser un impedimento, especialmente para pequeñas y medianas empresas. Por eso el papel catalítico de los apoyos públicos para que las empresas adopten esta tecnología es crucial.

Es bueno tener presentes los cinco principios de la IA industrial que Lee (2020) llama las cinco S: Sistemática, en el sentido de que la arquitectura debe llevar a una integración plena de la IA con los sistemas industriales; Speed (velocidad) para tener la posibilidad de diseñar, modelar, validar y desplegar sistemas de manera rápida para resolver problemas críticos y responder a necesidades del mercado; Streamline (simplificar) para generar procesos que aseguren resultados y se integren a la industria de forma simple y efectiva; Standards, para estandarizar datos, análisis, formatos y operaciones de la industria; y Sustainable (sustentable) para alcanzar un desempeño consistente y éxitos repetidos.

BIBLIOGRAFÍA

Abobakr. (2019). RGB-D ergonomic assessment system of adopted working postures. Appl. Ergonom., 75–88.

Asociación de Internet MX y Knowsy AI (2023). 19o Estudio sobre los hábitos de usuarios de internet en México 2023. Mayo. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://irp.cdn-website.com/81280eda/files/uploaded/19%20Estudio%20sobre%20los%20Hai-bitos%20de%20Usuarios%20de%20Internet%20en%20Mei-xico%202023%20.pptx.pdf

Bergamini. (2020). Deep learning-based method for vision-guided robotic grasping of unknown objects. Engonomic Informat, 13423.

Centro Nacional de Inteligencia Artificial (2024). Índice Latinoamericano de Inteligencia Artificial. https://indicelatam.cl/wp-content/uploads/2024/11/Ficha-Mexico.pdf

CISCO (2024). CISCO AI Readdiness Index. Hype Meets Reality. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.cisco.com/c/dam/m/en_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index.pdf

Chien. (2020). Deep reinforcement learning for selecting demand forecast models to empower industry 3.5 and anempirical study for a semiconductor component distributor. International Journal Production Research, 2784–2804.

Ibarra, Ernesto; De la Peña, Sissi; y Santoyo, Cristina; Panorama de la Inteligencia Artificial en México: hacia una Estrategia Nacional. 2024, disponible en: https://www.amcid.org/page/sandboxregulatoriomexico

IBM (2022). Global AI adoption Index 2022. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.eticanews.it/wp-content/uploads/2022/05/IBM-Global-AI-Adoption-Index-2022_FINAL.pdf

Instituto Mexicano para la Competitividad -IMCO- (2023). Inteligencia artificial: el futuro de la educación en México, Instituto Mexicano para la Competitividad, A.C., Ciudad de México

Lagunes A., Martínez Y., Cárdenas C., De la Peña, S., Mancilla, D., Xilotl R., Sánchez O., Moguel, A., Cárdenas J., (2024). "PROPUESTA DE AGENDA NACIONAL DE LA INTELIGENCIA ARTIFICIAL PARA MÉXICO (2024 - 2030)". Alianza Nacional de Inteligencia Artificial (ANIA). Disponible en: https://www.ania.org.mx/

Lai. (2020). Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. Journal Manufactur System, 69–81.

Lee, J. (2020). Industrial AI. Applications with Sustainable Performance, Springer, doi.org/10.1007/978-981-15-2144-7

Liang. (2019). Fog computing and convolutional neural network enabled prognosis for machining process optimization convolutional neural network enabled prognosis for machining process. Manuf. Syst, 32–42.

Liu, Q. (2024). Evaluating the Impact of 3D Printing Technology Innovations on Industrial Production Efficiency and Economic Growth in the Automotive Sector in China. Innovation Science and Technology.

May, C. y Rojas, D. (2024). Encuentros IA. Hoja de ruta para México. EON Institute y British Embassy in Mexico, Ciudad de México

Ojer. (2020). Real-time automatic optical system to assist operators in the assembling of electronic component. International Journal Manufactury Technology, 2261–2275. doi:10.1007/s00170-020-05125-z

Oldemeyer, L., Jede, A., Teuteberg, F. (2024). Investigation of artificial intelligence inSMEs: a systematic review of the state of the art and the main implementation challenges. Management Review Quarterly. https://doi.org/10.1007/s11301-024-00405-4

Oran. (2021). AI - Robotic Applications in Logistics Industry and Savings Calculation. Journal Of Organizational Behavior Research. doi:10.51847/juxqmvcvqf

Peres. (2019). Multistage quality control using machine learning in the automotive industry. IEEE Access, 79908–79916.

Piedras, E. (2024). Presupuesto TIC en la Primera Mitad de 2024: Asignaciones y Ejecución. The Competitive Intelligence Unit, Ciudad de México, 31 de octubre de 2024

Qin, J. (2018). Multi-so–850urce data analytics for AM energy consumption prediction. Adv. Eng. Informat.,, 840.

Schmitt. (2018). Implementation of cyber physical production systems for quality prediction and operation control in metal casting,. Sensors, 1428.

Schwab, K. (2017). The Fourth Industrial Revolution. Crown Publishing Group, New York.

Sections. (2021). Striving for excellence in AI implementation: AI maturity model framework and preliminary research results. 363–376. doi:https://doi.org/10.17270/J.LOG.2019.354

Solleiro, J. L., Castañón, R., Guillén, A. D., Hernández, T. Y., Solís, N. (2022). Boletín de vigilancia tecnológica en ciberseguridad 1. Instituto de Ciencias Aplicadas y Tecnología, UNAM. Descargable en https://www.icat.unam.mx/wp-content/uploads/2022/09/Vigilancia_Tecnologica_en_Ciberseguridad_Boletin.pdf

Solleiro, J.L. y Cuevas, J.L. (coord.) (2024). Guía para desarrollar una estrategia de ciberseguridad nacional, Instituto Federal de Telecomunicaciones, Ciudad de México

Szedlak, C., Leyendecker, B., Reinemann, H., Kschischo, M., Pötters, P. (2021). Risks and Benefits of Artificial Intelligence in Small and Medium Sized Enterprises. Proceedings of the International Conference on Industrial Engineering and Operations Management. Roma, Italua. Agosto 2-5. Disponible en chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ieomsociety.org/proceedings/2021rome/175.pdf

Yan. (2017). Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictiv maintenance,'. IEEE Access, 23484–23491.

IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN LA INDUSTRIA MANUFACTURERA. EL CASO DE MÉXICO

SUMARIO

I. Introducción **II.** La IA a lo largo de la cadena de valor de la industria manufacturera **III.** IA y la optimización de los niveles de inventario a lo largo de la cadena de producción **IV.** La IA en el proceso de fabricación del producto, **V.** Despacho y distribución del producto final **VI.** Capacidades para desarrollar la inteligencia artificial en la manufactura en México **VII.** Reflexiones finales **VIII.** Bibliografía.

> CLAUDIA SCHATAN

Especialista en temas económicos. Consejera del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI).

I. INTRODUCCIÓN

La industria manufacturera ha experimentado cambios radicales desde finales del siglo XX hasta los primeros años del siglo XXI, impulsados principalmente por el avance tecnológico. La digitalización y la integración de la tecnología de la información desarrolladas desde los años sesenta puede considerarse un pre-requisito para la introducción más reciente de la inteligencia artificial (IA) pues crearon la infraestructura para aplicarla, junto con el aprendizaje de las máquinas. Estas últimas tecnologías generan un valor agregado adicional a través de la extracción de información que con inteligencia computarizada puede accionar un proceso productivo con una optimización nunca antes lograda. El desarrollo en el área de IA se aceleró con la llegada de tecnologías avanzadas de computación y análisis de datos, especialmente a partir de la década de 2010, cuando se hizo posible el uso de macrodatos (o *big data*) e internet de las cosas (IoT)[1], computación en la nube, entre otras tecnologías de frontera, dando lugar a la "Industria 4.0" o manufactura inteligente, término acuñado en Alemania.[2] La posibilidad de

1 El Internet de las Cosas (IoT, por sus siglas en inglés) se refiere a una red de objetos físicos ("cosas") que tienen incorporados sensores, software y otras tecnologías con el fin de conectarse e intercambiar datos con otros dispositivos y sistemas a través de Internet (pueden utilizarse en una amplia gama de productos, desde bienes electrónicos digitales de consumo, hasta el control inteligente de ciudades. El Internet Industrial de las Cosas (IIoT, por sus siglas en inglés) se enfoca en la automatización y optimización de procesos industriales, únicamente.

2 Término acuñado por primera vez en la Feria de Hannover, Alemania en 2011. El concepto de Cuarta Revolución Industrial, similar a la Industria 4.0, lo introdujo Klaus Schwab (fundador del Foro Económico Mundial) en 2016 (https://es.wikipedia.org/wiki/Cuarta_Revolución_Industrial).

traducir información de procesos físicos en formato digital permite a las empresas una gran flexibilidad y optimización de su operación en tiempo real (Abd El Hedi Gabsi, 2024).

La IA, con su evolución hacia la inteligencia artificial generativa[3] (IAG) potencia las posibilidades de la industria en cada una de sus etapas de operación.

La manufactura inteligente recoge información mediante sensores colocados en las máquinas y otras partes del proceso de fabricación la cual se canaliza a través del IoT, y se procesa a través de algoritmos utilizando la IA. Ello da autonomía de decisión a sus propias máquinas y robots, reduciendo la necesidad de la intervención humana en una parte importante del proceso productivo, a la vez que economiza energía justamente por su mayor eficiencia, aunque, a su vez, indirectamente se requiere un gran consumo de energía.[4]

Con la información registrada por los sensores a lo largo de la cadena productiva –para lo cual se requiere de una muy eficiente interconectividad (con redes 5G)– se puede detectar y actuar con gran precisión e incluso antelación ante muy diferentes situaciones. Es posible prever la necesidad de hacer mantenimiento a las máquinas; evitar accidentes; ubicar defectos en el producto y corregirlos; percibir cuellos de botella, entre otros, pudiendo así actuar en forma automática y descentralizada, lo que contribuye a mejorar notablemente la productividad. A través de la digitalización y la IA, también es posible percibir los cambios en las características de la demanda de los productos en el mercado, cada vez más personalizada, lo que ayuda a incorporar en forma automática las preferencias de los usuarios en los procesos productivos.

La IA en la manufactura no se limita al ámbito de las unidades productivas de una empresa sino a toda la cadena de valor. Es decir, la IA interviene desde la etapa del diseño de un producto, el manejo de los inventarios, el proceso

3 Inteligencia Artificial Generativa (IAG): se utilizan modelos extensos de lenguaje (llms, por sus siglas en inglés) para textos, y modelos de difusión de imágenes. Estos últimos son más profundos (es decir, tienen más capas de neuronas) que los anteriores, y están organizados de manera que les permiten moverse rápidamente a través de gran cantidad de datos, The Economist, https://www.economist.com/schools-brief/2024/08/06/how-ai-models-are-getting-smarter).

4 La IA ayuda a optimizar procesos industriales y, por tanto, reduce el consumo de energía. Por ejemplo, la optimización de procesos productivos, a través de algoritmos puede ajustarlos continuamente para minimizar el uso de energía apoyándose en el uso de información transmitida en tiempo real. Por otra parte, la IA requiere de un apoyo externo a la empresa, por ejemplo, de los centros de datos, que son crecientes consumidores de energía, en la medida en que se expande el uso de la IA. El entrenamiento de la IA requiere grandes cantidades de energía y cada vez más, cuanto más sofisticada la tecnología.

de transformación, la logística de transporte de insumos y producto final, los procesos de comercialización y distribución de productos. La intervención de la IA para reducir cuellos de botella en las cadenas y eliminar tiempos muertos por insuficiencia de insumos, por ejemplo, agiliza mucho el proceso productivo.

En este ensayo exploraremos, en primer lugar, las transformaciones que está experimentando el sector manufacturero de punta con la IA, cuáles son las tecnologías específicas que está asimilando esta industria a lo largo de su cadena de valor, así como sus necesidades de energía y capital humano. En segundo lugar se explora el panorama para la incorporación de la IA en la industria manufacturera en México y cuales son algunas precondiciones para un mejor aprovechamiento de esta tecnología. Por último presentaremos algunas reflexiones finales.

II. LA IA A LO LARGO DE LA CADENA DE VALOR DE LA INDUSTRIA MANUFACTURERA:

A. IA EN EL DISEÑO DEL PRODUCTO

La IA puede dar lugar a un profundo cambio de parámetros en el diseño de los productos manufacturados. Al inicio de la cadena de valor de un determinado producto manufacturado se encuentra la etapa de diseño en la cual la IA puede ser de gran utilidad. En primera instancia, a través de algoritmos apropiados, pueden analizarse grandes volúmenes de datos de diseño para identificar patrones y proponer mejoras de diseños ya existentes. Al mismo tiempo, a través del análisis de datos y aprendizaje automático, la IA puede predecir las preferencias del consumidor y adaptar los diseños de acuerdo a ellas. Incluso, con algoritmos generativos de IA, se pueden crear múltiples versiones de un diseño basándose en criterios predefinidos (peso, resistencia, costo, etc.), lo que permite, dada la tendencia que existe a personalizar masivamente los productos, responder a las necesidades específicas de cada cliente.

En segundo lugar, una vez que se cuenta con el diseño del producto, las empresas que necesitan elaborar prototipos de éste antes de proceder a su fabricación, pueden utilizar la IA para la creación y la evaluación virtual de dichos prototipos antes de incurrir en el costo de construir un modelo físico. Ello da gran flexibilidad y ayuda a optimizar el diseño, al permitir efectuar múltiples iteraciones de éste de manera rápida y eficiente hasta llegar a una versión óptima en forma segura y fiable.

Un prototipo concebido de manera virtual da oportunidad a los diseñadores de observarlo a través de la realidad virtual (RV) en tres dimensiones dentro de un entorno inmersivo. Esto es especialmente útil para revisar detalles complejos del diseño que podrían no apreciarse en representaciones en 2D o en pantallas planas. En el mundo virtual, los diseñadores explorarán el producto en 3D desde todos los ángulos y realizarán los cambios que les parezcan necesarios. Además, varios diseñadores pueden colaborar en su concepción desde distintos lugares geográficos sin tener que estar físicamente presentes en la fábrica (Berg y Vance, 2017).

Finalmente, y en congruencia con lo anterior, en este mundo virtual (RV) se pueden probar los productos en diferentes escenarios y condiciones únicamente con el diseño antes de ser fabricados físicamente y pueden hacerlo incluso los potenciales compradores. Esto facilita a los diseñadores y a los potenciales usuarios del producto final observar cómo interactúan las personas con el producto, identificar posibles problemas de ergonomía o funcionalidad, por ejemplo, y realizar ajustes antes de la fabricación.

El desarrollo de productos de manera digital también puede implementarse mediante gemelos digitales, que son réplicas virtuales de productos existentes[5] que simulan todas las características de sus contrapartes físicas. El interactuar con estos gemelos virtuales puede ayudar a modificar un producto ya existente pero en un espacio virtual en forma mucho más fácil, rápida, y más segura que en el mundo físico. Esta tecnología también ha demostrado ser muy útil en el diseño de productos, entre otros usos.

B. INTELIGENCIA ARTIFICIAL EN LA FASE DE PRODUCCIÓN

Una vez diseñado el producto como descrito en la sección anterior, la IA puede tener un rol crucial para mejorar u optimizar el funcionamiento del resto de la cadena de valor de la industria a través de varias vías. Se puede optimizar la Cadena de Suministro para el conjunto del proceso productivo a través de la IA (con base en datos históricos y otras tendencias específicas) al predecir desde un inicio cuál será la demanda del producto, lo cual orienta a las empresas a ajustar, de acuerdo a ello, el inventario en las distintas etapas de la cadena e intervenir en su proceso de producción.

5 Los gemelos digitales pueden aplicarse a productos, procesos manufactureros o incluso a una cadena de suministro completa.

III. IA Y LA OPTIMIZACIÓN DE LOS NIVELES DE INVENTARIO A LO LARGO DE LA CADENA DE PRODUCCIÓN

La IA contribuye a hacer un análisis predictivo de la demanda que habrá por el producto elaborado y, de acuerdo a ello, los inventarios que se requerirán a lo largo de la cadena de producción. Es decir, un algoritmo de IA y, aún más uno de IAG, es capaz de analizar los datos históricos de ventas, considerando estacionalidad, tendencias del mercado, entre otras variables, lo que ayudará a la empresa a mantener los stocks de inventarios necesarios para garantizar una producción continua, sin que haya exceso o falta de insumos y partes.

Se puede, asimismo, programar el reabastecimiento automático de inventarios, ordenando la cantidad óptima de insumos o partes de acuerdo a la rotación que tienen dichos stocks, para lo cual ayuda enormemente el IoT y *Big Data* junto con la IA.

Lo anterior es congruente con el concepto de Planificación de Producción Justo a Tiempo (Just-in-Time, JIT), adoptado ampliamente en las cadenas de valor de la producción internacional hasta antes de la pandemia de 2019-2022, aunque fue re-evaluado debido a que, entonces, la escasez de inventarios estranguló muchos procesos productivos. Pero la IA puede ayudar a indicar la necesidad de inventarios, y sus lugares de almacenamiento, considerando mayor holgura en la disponibilidad de ellos en las cadenas con el fin de evitar situaciones de gran escasez de insumos ante cuadros de pandemias o de eventos climáticos extremos. En principio, la IA puede ajustar dinámicamente los niveles de inventario en respuesta a cambios en tiempo real, como variaciones inesperadas en la demanda o problemas en la cadena de suministro.

En síntesis, la IA ayuda a optimizar los niveles de inventario a través de una combinación de análisis predictivo, automatización del reabastecimiento, optimización de la producción, y minimización de residuos. Se logra, así, una mayor eficiencia, la reducción de costos y la respuesta rápida de la cadena de valor a la demanda del mercado.

Más allá de lo mencionado, la IA puede contribuir a ubicar diversos proveedores de suministro de insumos, así como comparar diversas ofertas por diferentes empresas y manejar y escoger las mejores opciones para licitar, acorde con diferentes escenarios. Asimismo, puede ayudar a llevar a cabo los pagos necesarios cumpliendo con los diversos pasos bancarios y tributarios en forma automatizada.

Finalmente, al manejo de inventarios con IA, hay que agregar que la eficiencia mejora aún más con la robótica y la automatización guiada por la IA. Por ejemplo, los robots autónomos solos o en colaboración con trabajadores pueden ser usados en los almacenes en que se guardan los inventarios, para escoger, empacar y organizar los stocks en forma eficiente.

IV. LA IA EN EL PROCESO DE FABRICACIÓN DEL PRODUCTO

La eficiencia del proceso de transformación manufacturera puede elevarse considerablemente al usar instrumentos de IA para entrenar robots en el proceso de producción, así como para hacer un mantenimiento predictivo de la maquinaria y la supervisión de la calidad del producto, con posibilidades de corregir en forma inmediata las irregularidades que puedan detectarse en estos frentes (Deloitte, 2023). La IA en la manufactura abarca tanto el aprendizaje automático[6] como el aprendizaje profundo.[7]

ENTRENAMIENTO DE ROBOTS PARA LA PRODUCCIÓN MANUFACTURERA

La IA puede usarse para activar robots en el proceso de transformación productiva, en el que se analiza información para orientar y mejorar el desempeño productivo. Los robots articulados son los más usados actualmente en el mundo industrial. Tienen una estructura mecánica similar a la de un brazo humano, el cual se monta en una base de articulación rotatoria. Los algoritmos de la IA ayudan a estos robots a aprender a controlar sus movimientos y manejar los objetos en forma precisa. La información recogida y analizada en tiempo real con IA puede ser usada para informar y guiar las acciones de estos robots. Con la IA se

6 Aprendizaje automático: es "una rama de la inteligencia artificial que permite que las máquinas aprendan ciertas tareas sin ser programadas de manera específica para ellas. Para ello, utilizan estadísticas para predecir y reconocer patrones, por ello funcionan mejor en grandes conjuntos de datos." (https://www.bbva.com/es/innovacion/machine-learning-que-es-y-como-funciona/#).

7 Aprendizaje profundo: se basa en capas de las redes neuronales, que son algoritmos que asemejan el funcionamiento del cerebro humano. Este entrenamiento requiere de grandes cantidades de datos y logran mejor desempeño en la medida en que absorben más información a lo largo del tiempo.

crean sistemas manufactureros más flexibles y adaptables, capaces de manejar numerosas tareas simultáneamente y cada vez con mayor exactitud.

En la industria automotriz, por ejemplo, se usan robots en casi todas las etapas de la producción, desde la fabricación de partes hasta el ensamblaje final y se utiliza la IA para monitorear y mejorar el trabajo de los robots continuamente. La automotriz Tesla, por ejemplo, cuenta con "Gigafábricas" para automatizar el proceso de producción. Los robots controlados por IA realizan tareas como el ensamblaje de vehículos, la soldadura y la pintura con alta precisión. Estos robots están equipados con sensores y algoritmos que les permiten aprender y adaptarse para mejorar la eficiencia y la calidad de producción.[8]

MANTENIMIENTO PREDICTIVO

En cualquier fábrica, es indispensable dar mantenimiento a la maquinaria. Tradicionalmente, los enfoques respecto de esta actividad incluyen (i) "descompostura y arreglo", que es compostura correctiva (CM, por sus siglas en inglés) y consiste en arreglar el equipo sólo después de haberse dañado y responde sobre todo a arreglos que no presentan mayores riesgos. Esta forma de reparar equipo puede significar costos significativos, incluyendo los tiempos muertos de la maquinaria mientras se realiza la reparación; (ii) cuando es necesario apegarse a estándares de seguridad más severos porque hay más riesgos involucrados, el mantenimiento preventivo (PM, por sus siglas en inglés) es una estrategia más apropiada que la anterior, pues evita la ocurrencia de fallas catastróficas para el equipo industrial. Esta contempla una revisión periódica, de acuerdo a la información que se tiene sobre la durabilidad de la maquinaria y sus piezas (esto se puede conocer desde el diseño del producto). Este último enfoque de todas maneras puede resultar no óptimo porque requiere la interrupción del proceso productivo mientras se efectúa y podrían cambiarse piezas que posiblemente aún estén en buenas condiciones, además de desperdiciar materiales, entre otras limitaciones (Ruiz-Sarmiento, *et al*, 2024; Chao Huan, *et al*, 2024).

Alternativamente, usando IA, el mantenimiento predictivo (PdM, por sus siglas en inglés) está específicamente diseñado para orientar la acción de mantenimiento sobre la base de las condiciones de operación en tiempo real y

8 Tesla Mag (03/09/2024), https://www.tesla-mag.com/en/tesla-gigafactories-pioneering-the-future-of-sustainable-manufacturing/#:~:text=Advanced%20Automation%20and%20Artificial%20Intelligence,faster%20and%20more%20precise%20manufacturing.

las proyecciones que se tienen. El PdM es, por tanto, una práctica mucho más eficiente para conservar la planta en óptimas condiciones y alcanzar una máxima productividad. Haciendo uso de numerosos sensores que se colocan sobre la maquinaria y otros sitios de las instalaciones, se transmite información valiosa (temperatura, presión, humedad, vibración, etc.) a través del Internet de las cosas (IoT) en tiempo real, que se procesa por medio de algoritmos. El resultado de este proceso es una predicción oportuna respecto del momento en que podría ocurrir una falla, facilitando la decisión acerca del mantenimiento que es necesario realizar en el momento justo y reduciendo así el tiempo y el costo de estos procedimientos. Es más, al predecir un problema, el sistema de IA en la empresa puede recomendar una acción o tomar una decisión inteligente en forma automática para que se supere el problema incluso sin la intervención humana. Los mejores algoritmos –existen una gran variedad de ellos– para tomar las mejores decisiones en los momentos más adecuados son materia de discusión (Bousdekis, *et al*, 2019). El PdM también contribuye a mejorar la seguridad en la planta, con menos accidentes, así como a prolongar la vida de la maquinaria y a mejorar la calidad de los productos (pues el monitoreo reduce el número de productos defectuosos). En el caso del PdM el uso de gemelos digitales, al igual que en la etapa de diseño del producto, también pueden ser de gran utilidad, especialmente a medida que se generaliza el IoT. La realidad aumentada (RA)[9], por su parte, que es una combinación de realidad virtual y física es sumamente útil para la compostura y mantenimiento de la maquinaria que se usa en una fábrica, pero también, por ejemplo, para el mantenimiento y reparación de los vehículos en los talleres de servicio. Es más, los técnicos o ingenieros pueden acceder remotamente a través de un dispositivo sensorial como lentes de RA para hacer diagnósticos y dirigir la reparación física de la maquinaria o vehículos desde lugares remotos.

Una visión más amplia del mantenimiento de la planta, que incluye el PdM es el de Pronósticos y Gestión de la Salud (PHM, por sus siglas en inglés), que es un "enfoque de mantenimiento de máquinas que utiliza datos de sensores históricos y en tiempo real para fundamentar y optimizar decisiones de mantenimiento. El PHM es un enfoque integrado que combina dos conceptos clave: '*Pronósticos*: Enfoque de diseño de algoritmos para estimar la vida útil restante de sistemas o componentes. Este término a menudo se emplea como sinónimo de *mantenimiento predictivo*' y '*Gestión de la salud*: Enfoque de

9 La realidad aumentada añade elementos digitales a un entorno real, pero no es una inmersión, sino es una interacción entre lo digital y lo real.

mantenimiento integral que aplica la información de algoritmos de pronósticos y diagnósticos, entre otros, para garantizar la salud y fiabilidad del sistema.'"[10]

CONTROL DE CALIDAD

Una vez que el producto manufacturado está concluido, se pasa a la etapa de inspección de su calidad y, de acuerdo a la evaluación que se haga, el producto se acepta o se rechaza. Asegurar la calidad del producto es esencial para garantizar la seguridad y salud del consumidor, así como su satisfacción con el producto que está adquiriendo. Para esta etapa se necesita una inspección visual, pero ésta a simple vista humana tiene una precisión del 80% en la industria y lograr el 100% es muy costoso, tardado y difícil. Las cámaras de alta resolución junto con algoritmos integrados (Computer Vision, CV) han ayudado a automatizar partes del proceso de inspección visual, agiliza esta etapa y permite identificar los defectos con gran precisión (Sundaran y Zeid, 2023). Nótese que la tecnología de visión computarizada también es sumamente útil para el mantenimiento predictivo antes descrito.

Cuanto más sofisticados tecnológicamente los productos de los que se trate, más importante es el CV. En la industria de semiconductores, por ejemplo, versiones muy refinadas de esta tecnología CV junto con el aprendizaje automático de máquinas detectan defectos erráticos en imágenes microscópicas de electrones y mapas de obleas de silicio, lo que resulta muy útil para poder asegurar el correcto funcionamiento de un microchip (Plathottam, *et al*, 2023).

V. DESPACHO Y DISTRIBUCIÓN DEL PRODUCTO FINAL

En esta etapa del ciclo del producto, la IA puede hacer una importante contribución para visibilizar el tiempo de entrega y optimizar la ruta de distribución (Deloitte, 2023). El acceso inmediato y detallado de información en la etapa final de la cadena de valor es esencial para reducir los riesgos de distribución y aumentar la eficiencia en entrega de los productos finales.

La IA puede optimizar la logística, las rutas de transporte y la gestión de flotas, reduciendo tiempos de entrega y costos operativos, mientras se minimiza la huella de carbono. Tradicionalmente, la logística y transporte de

10 Mathworks, *Pronósticos y Gestión de la Salud* (PHM), https://la.mathworks.com/discovery/prognostics-and-health-management.html.

los productos se ha realizado manualmente, pero el uso de la IA puede indicar las vías más eficientes de distribución mediante algoritmos sofisticados para analizar diversos factores desde los puntos de despacho, los patrones de tráfico por distintas rutas, las condiciones de carreteras u otras vías de transporte, y el consumo de combustibles. Una vez en ruta, se rastrean los vehículos de transporte utilizando sensores colocados en ellos lo que permite a la industria recabar información en tiempo real a través del IoT sobre su ubicación, su velocidad y dirección, de manera que se puede responder en forma inmediata a disrupciones con un ajuste en las rutas originalmente escogidas, evitando o reduciendo retrasos u obstáculos en la fase de distribución y entrega de los productos. Adicionalmente, el análisis de mantenimiento (PdM) predictivo también ayuda en este caso a mantener la flota de transporte en las mejores condiciones a través de su monitoreo mediante sensores en tiempo real y puede realizarse el mantenimiento anticipadamente, evitando desperfectos.

La ruta emprendida por los productos en su fase final se puede agilizar considerablemente también con la automatización del procesamiento de los documentos, como parte de una logística digitalizada. Los sistemas de IA para este fin utilizan programas como Reconocimiento Óptico de Caracteres (OCR, por sus siglas en inglés) y Procesamiento Natural de Lenguaje (NLP, por sus siglas en inglés) para leer, comprender, y obtener información esencial desde documentos físicos y digitales, lo que facilita el procesamiento y la transmisión de la información, reduciendo así errores manuales y mejorando la eficiencia operacional de estos procesos.

Amazon, líder en el comercio electrónico mundial, puede ejemplificar la logística en la distribución final de productos utilizando AI. Sus algoritmos ayudan en la previsión de la demanda de productos, el manejo óptimo de stocks, sus modelos de almacenaje automatizados (en los que los robots realizan mucho del acomodo y empacado de las mercancías en sus bodegas). Fedex también utiliza IA para automatizar y optimizar la clasificación de sus paquetes en sus centros de acopio. Sus robots activados con inteligencia artificial pueden clasificar hasta 1,200 paquetes por hora. También usa IA para la predicción analítica de potenciales retrasos a consecuencia de elementos disruptivos imprevistos en sus entregas, pudiendo así alertar a sus clientes.[11]

11 Eliftech (11/01/2024), *AI in Logistics: How Does It Truly Transform The Field?*, https://www.eliftech.com/insights/ai-in-logistics-explained/

A. LA IA, EL USO DE ENERGÍA Y LAS EMISIONES DE LA INDUSTRIA MANUFACTURERA

Varios de los instrumentos de IA utilizados en la cadena productiva de la industria manufacturera ya descritos también son de gran ayuda para racionalizar el uso de energía y reducir las emisiones de la industria manufacturera. Por ejemplo, el monitoreo y mantenimiento predictivo de maquinaria y equipo, ayuda a evitar fallas que generan emisiones no deseadas o bien desperdicio de energía. Con sensores y herramientas de visión artificial se pueden incluso detectar emisiones específicas, como fugas de metano, que pueden ser atendidas rápidamente, evitando su impacto ambiental.

Al aplicar IA y el Internet de las Cosas (IoT), se puede avanzar hacia una economía circular. La IA ayuda a rastrear y reducir las emisiones de la cadena de suministro mediante el análisis de datos de cada etapa del proceso de producción. Al identificar los puntos críticos de emisión, es posible implementar mejoras específicas, además de gestionar el reciclaje y reutilización de materiales, disminuyendo la necesidad de materias primas nuevas y reduciendo las emisiones indirectas.

Otro mecanismo mencionado anteriormente – el de gemelos digitales - puede ser una gran ayuda en el ahorro de energía en la manufactura. Estos crean simulaciones virtuales de sistemas físicos, lo que permite a las empresas probar y ajustar procesos antes de implementarlos en la realidad, y con ello optimizan el uso de recursos energéticos sin tener que realizar cambios físicos, que pueden ser costosos y menos sostenibles. Estos modelos de IA facilitan una supervisión constante y ajustes de consumo energético basados en datos, generando ahorros considerables y reduciendo el impacto ambiental.

Además, los sistemas de IA son capaces de analizar datos en tiempo real para ajustar y optimizar el uso de energía en las líneas de producción. Por ejemplo, la integración de IA con sensores IoT permite ajustar el consumo de energía en función de la carga de trabajo y las demandas del momento. Esto reduce el consumo en períodos de menor actividad o durante picos de energía renovable, lo que mejora tanto el costo como la sostenibilidad de las operaciones.

B. LA REVOLUCIÓN DE LA CAPACITACIÓN DEL CAPITAL HUMANO EN LA INDUSTRIA MANUFACTURERA CON IA

La IA en la industria manufacturera es muy útil para entrenar a los trabajadores en múltiples tipos de tareas y para ello son especialmente apropiadas las tecnologías de realidad virtual inmersiva (RV) y realidad aumentada (RA).

La RV permite simular digitalmente el espacio de trabajo en forma virtual en 360° reproduciendo en esa dimensión los equipos y herramientas para que un trabajador se pueda desenvolver plenamente en este plano imaginario. Esta tecnología permite a los trabajadores experimentar simulaciones de procesos industriales complicados sin riesgos, lo que es especialmente útil para entrenar en operaciones peligrosas o complejas, como el manejo de maquinaria pesada o líneas de producción automatizadas. Por ejemplo, los empleados pueden practicar el ensamblaje de piezas en ese entorno antes de enfrentarse a la maquinaria real, lo que reduce errores y mejora la seguridad. Los conocimientos adquiridos en forma virtual se aplican al mundo real una vez concluido tal entrenamiento.

Con la RA, por otra parte, los técnicos pueden ver instrucciones paso a paso superpuestas sobre las máquinas reales (realidad mixta) a través de gafas o dispositivos móviles. Esta visión puede ser compartida entre especialistas ubicados en distintos puntos geográficos y abre la posibilidad de interactuar a distancia para encontrar soluciones a problemas observados. Para el entrenamiento de trabajadores es especialmente útil porque éstos pueden recibir la instrucción vía remota por lo que la empresa puede tener muy buenos maestros a distancia. Empresas como Boeing y Siemens, por ejemplo, utilizan la RA para guiar y entrenar empleados en la reparación de equipos complejos. (Medium, 2021; Siemens, 2024).

Con la RV, los empleados de distintas plantas pueden entrenarse juntos en un entorno virtual común, mientras que la RA permite a expertos supervisar, entrenar y guiar a distancia a los trabajadores en tiempo real. En ambos casos, los tiempos y los costos de la capacitación se reducen considerablemente respecto a los métodos tradicionales. En Boeing, por ejemplo, se estima que se ha reducido en 75% el tiempo de entrenamiento gracias a la RA (Medium, 2021).

En pocas palabras, no hay prácticamente ninguna fase de la cadena de valor de la manufactura en la cual no se pueda aplicar la IA con resultados de aumento de la productividad.

VI. CAPACIDADES PARA DESARROLLAR LA INTELIGENCIA ARTIFICIAL EN LA MANUFACTURA EN MÉXICO

La introducción de la IA en México es ya una realidad en diversas esferas, aunque incipiente aún, si se consideran todas las potenciales aplicaciones en la cadena de valor descritas anteriormente. En el sector manufacturero mexicano la IA se está incorporando especialmente a las empresas globales.[12]

Las condiciones para el desarrollo de la IA en México, si bien han avanzado, tienen que reforzarse, particularmente si pretende convertirse en un polo atractivo para el *nearshoring* con Estados Unidos pues muchas de las empresas que tienen intenciones o pueden sentirse atraídas para instalarse en el país necesitarían incorporar la IA en sus cadenas de valor.

Para conocer el desempeño de México en el ámbito de la IA o su grado de preparación para incorporarla, en general, utilizaremos algunos indicadores que, si bien no nos dan una visión específica de la IA en el sector manufacturero, pueden ayudar a tener un punto de referencia para este último.

Es interesante observar, en primer lugar, la capacidad innovadora de México, que se liga crecientemente con sus posibilidades de asimilar la IA. El Indice de Innovación Global (GII, por sus siglas en inglés) nos señala que México ocupa el lugar 56 entre 133 economías en el mundo y el tercer lugar entre 20 países latinoamericanos en 2024 (WIPO, 2024). Si bien México aparece teniendo un desempeño en innovación acorde con su nivel de desarrollo (PIB per cápita, PPP en escala logarítmica), su potencial podría ser mayor si se le compara con países como Brasil, India, Vietnam y especialmente China, que tienen un desempeño en innovación por encima del que se esperaría considerando sus PIB per cápita (WIPO, 2024). Según esta misma fuente, la tendencia no es alentadora para México pues la inversión en ciencia e innovación ha ido a la baja en el país, particularmente en los últimos años, incluyendo publicaciones científicas, inversiones en I&D, inversiones en capital de riesgo y en solicitud de patentes.

Al observar, más específicamente, el Indice de Inteligencia Artificial Global, que se apoya en 122 indicadores y clasifica 83 países a nivel mundial en materia de capacidad en IA, vemos que los punteros internacionales en este indicador

12 La implementación de IA en manufactura implica costos elevados en equipos, software y adaptación de procesos. Las pequeñas y medianas empresas (PYMES), que representan una parte significativa del sector manufacturero mexicano, suelen carecer del capital para invertir en soluciones avanzadas. Esto se ve agravado por la falta de programas de financiamiento específicos para la digitalización industrial.

son Estados Unidos en primer lugar, siguiéndole China, mientras que México ocupa el lugar 45, lo cual indica un reto para que el país pueda ser competitivo en este ámbito (Cuadro 1). Brasil se encuentra en el lugar 30, bastante más arriba que México. Nótese que los aspectos más débiles de la preparación de México para la IA están la investigación en IA (por parte de la academia y de la industria) (lugar 61); en infraestructura (de informática avanzada) y la producción de microchips (lugar 57). También requiere mejorar la estrategia gubernamental (lugar 47) y contar con mayor talento (lugar 42) con el fin de estar en mejores condiciones para la adopción de IA. En el entorno operativo, el país tiene un mejor ranking (20), lo que indica un marco regulatorio favorable [13] y una actitud positiva de la sociedad hacia la IA. Cabe destacar que México está mejor situado que Vietnam (lugar 58) –competidor cercano a México en el *friend-shoring*– en la mayoría de los indicadores escogidos.

CUADRO 1 Indice de Inteligencia Artificial Global

Indicadores escogidos (capacidad para el desarrollo de la IA)*

País	Ranking	Talento	Infra-estructura	Entorno Operativo	Inves-tigación	Desarrollo	Estrategia Gob.
USA	1	1	1	2	1	1	2
China	2	9	2	21	2	2	5
Rep. Corea	6	13	6	35	13	3	4
España	18	18	19	17	26	21	6
Brasil	30	26	36	28	44	29	27
México	45	42	57	20	61	40	47
Vietnam	58	49	33	70	67	58	56
Perú	61	61	52	37	81	75	60

FUENTE: Tortoise Media, The Global AI Index 2024, https://www.tortoisemedia.com/intelligence/global-ai/

13 Aunque el país no tiene aún una Estrategia para la IA oficial y hay muchas lagunas legales aún.

*** INDICADORES ESCOGIDOS:**

Talento: se enfoca en la disponibilidad de profesionales capacitados en soluciones de inteligencia artificial.
Infraestructura: evalúa la disponibilidad de infraestructura informática avanzada y la fabricación de semiconductores.
Entorno operativo: se centra en el contexto regulatorio y la opinión pública sobre la inteligencia artificial.
Investigación: analiza la cantidad y la calidad de la investigación académica y de la industria especializada.
Desarrollo: se centra en el desarrollo de nuevos modelos de IA y en la aplicación de la tecnología de IA en patentes en otros ámbitos.
Estrategia Gubernamental: mide el grado de compromiso de los gobiernos nacionales con la inteligencia artificial; sopesa el gasto y las estrategias nacionales.

Al comparar a México con otros países de la región latinoamericana, a través del Indice Latinoamericano de Inteligencia Artificial de la Comisión Económica para América Latina y El Caribe (CEPAL, 2024) que analiza la situación de 19 países respecto a la IA, México se ubica en el 6° lugar, con un puntaje de 51.4 (el primer lugar lo tiene Chile con un puntaje de 73.1 y en el otro extremo está Honduras con un puntaje de 23.7). De acuerdo a este estudio, México es un país con un "ecosistema maduro y sólido" dentro de una región que se encuentra en un período de "arranque". El punto de referencia de América Latina es poco indicativo de la ubicación de México en el mundo en cuanto a preparación para asimilar la IA, dado que la región mantiene una fuerte brecha con los países del norte global. Por ejemplo, el puntaje que alcanzan, en promedio, los 19 países de ALC para los factores habilitantes de la IA (infraestructura, acceso a datos y talento humano, sin los que no puede haber asimilación de IA) es de 43.1 en infraestructura; 35.8 en accesibilidad a datos; 39.7 en talento humano y en ninguno de estos indicadores la región, ni ningún país específico llega al máximo puntaje de 100). México está algo por arriba de los puntajes promedio en factores habilitantes, pero no en forma muy significativa: 51.0, 48.2, y 44.9, respectivamente. En el indicador de conectividad, México tiene el quinto lugar en América Latina, sin embargo el tipo de conectividad que se necesita para el uso de IA (especialmente 5G) requiere de un esfuerzo especial. Aunque México se encuentra bien ubicado en disponibilidad de 5G en el conjunto de los países de ALC escogidos por el indicador de CEPAL, no cuenta con lo necesario para un verdadero arranque en IA (como se mencionará más adelante).

Vale la pena destacar que el ILIA de 2024 (CEPAL, 2024) tiene un indicador sobre la incorporación de la IA en la Industria. Para construirlo, se usan tres subindicadores: trabajadores en sectores de alta tecnología; fabricación de tecnología mediana y alta; y proporción del valor añadido de fabricación de mediana y alta tecnología. México y Brasil aparecen con muy altos puntajes en este indicador: 96.3 y 83.8, respectivamente. Estos dos países tienen la industria manufacturera más grande, sólida y de tecnología más avanzada en la region de ALC y bien pueden ser los países en cuya manufactura penetre más fácilmente la IA, pero por ahora, se encuentra en estadios incipientes aún.

En síntesis, el informe ILIA de CEPAL (2024) indica que México necesita fortalecer la infraestructura, mejorar el talento humano en IA, impulsar la investigación, mejorar la institucionalidad y la gobernanza entre otras cosas.

A partir de ciertos elementos que destacan en los índices antes descritos analizaremos algunas condiciones que necesita México para que la industria manufactura pueda funcionar con IA en forma adecuada. Un primer punto es que la Industria 4.0 requiere de la IA para funcionar como una industria inteligente para lo cual es necesaria la disponibilidad de redes de quinta generación de las tecnología celular (5G) y México dista de contar con la infraestructura requerida (véase Cuadro 1 nuevamente). Las redes 4G, o su versión más avanzada, LTE, y el Wi-Fi (hasta ahora) son insuficientes para que la Industria 4.0 pueda operar plenamente. Las redes 5G son las que hacen posible la transmisión de información casi inmediata dentro de las empresas y con sus proveedores, es decir, ofrecen la conectividad esencial en la actual revolución tecnológica. Estas redes son las que dan las mayores velocidades de transmisión de datos, la menor latencia (o el tiempo que transcurre entre una solicitud de información a internet y su respuesta, que puede llegar a ser de un milisegundo) y el acceso a muchos más dispositivos que con las redes anteriores, lo que permite usar mucho más ampliamente el IoT. A lo anterior hay que agregar que la red 5G es más segura que sus antecesoras, pues la información se encripta en forma más sofisticada. Por último, es posible segmentar la red 5G de manera que los operadores pueden hacer uso de secciones propias, privatizadas, para fines específicos sobre una infraestructura de red común (por ejemplo, una red privada 5G puede usarse para aplicaciones de vigilancia a través de videos, o para dar soporte a los robots de una empresa, entre otros)(Schatan, 2023).

Si bien la introducción de 5G avanza en México, su pleno uso requerirá de muy importantes inversiones en infraestructura, por ejemplo, en gran número de torres y antenas, pues se necesitan más de ellas para cubrir la misma área de 4G porque las señales de 5G tienen un alcance más corto que las anteriores. Se

necesita, sobre todo, subastar una considerable cantidad de espectro 5G. Ello se ha estado planeando desde 2021 (licitación IFT-12), pero se ha postergado hasta 2025, si bien la consulta previa a tal licitación ya se realizó en 2024. Se está considerando ofrecer bloques de espectro en 600 MHz, 800 MHz, Banda L, Banda AWS, 2.5 GHz, entre otras (IFT, 2024). De todas formas, no está asegurado el éxito de esta licitación pues podría ocurrir, al igual que con la licitación IFT-10 (2021), que haya una limitada adquisición del nuevo espectro por parte de las empresas debido al alto precio a las que se ofrece (que en ese entonces fue 60% más cara que el promedio mundial). Incluso ha habido devoluciones de espectro por parte de empresas como Telefónica, por no resultarles redituable (Schatan, 2023). En la licitación de 2025 el IFT tiene planeado ofrecer importantes rebajas en el precio del espectro especialmente para aquellos que no tengan concesiones anteriores, pero la definición de todo ésto aun estaba en proceso a fines de 2024 (IFT, 2024).

Aún con los planes que se tienen, hay retos que no van a ser superados enteramente "…las redes de las empresas que actualmente ofrecen redes 5G hacen uso de partes de frecuencias reordenadas en las bandas ya existentes de 3.4 GHz y 2.5 GHz, que no proveen el ancho de banda que necesitan las nuevas redes 5G para operar con su máximo potencial. Es indispensable tener una mayor infraestructura en telecomunicaciones en México que ayude al despliegue de la red 5G en formato Stand Alone (SA), que es el que ofrece la velocidad de descarga máxima (requiere una infraestructura completamente nueva) en lugar de Non-Stand Alone (NSA) que funciona sobre la infraestructura de la red 4G LTE y ofrece velocidades inferiores." (Schatan 2023, pág. 33). Todo ello es crucial si se pretende que la IA pueda operar a lo largo de la cadena de valor de la industria manufacturera, como descrito en la sección 1.

Otra de las importantes limitaciones que enfrenta México para introducir la IA es que, aunque cuenta con talento en ingeniería, tiene una gran escasez de conocimiento especializado en IA, así como en ciencia de datos y otras áreas tecnológicas vinculadas a la IA. La incorporación de tecnología de punta, incluyendo la IA en el sector manufacturero, requerirá el mejoramiento de las habilidades existentes a través de la re-cualificación (*upskilling)* además del desarrollo de nuevas competencias (*reskilling*) del capital humano. La oferta de formación técnica y universitaria en IA es aún limitada en México, aunque se están introduciendo nuevas iniciativas y nuevos programas de formación vinculados a la IA, como señala un estudio reciente auspiciado por la Embajada de Gran Bretaña en México y que recaba información de varias fuentes (Ibarra, De la Peña y Santoyo, 2024). En este último se menciona la conformación reciente del Grupo de Inteligencia

Artificial (GIA) para la Educación Superior. Una consulta realizada por el GIA-IES a integrantes de la ANUIES para el ciclo educativo 2022-2023, refleja una limitada oferta oficial por parte de las instituciones de educación superior (IES) en cuanto a programas específicos para la formación y especialización en IA. Todos ellos son muy recientes y, por tanto, con muy pocos egresados aún en el mercado de trabajo: en el ciclo educativo del 2022-2023 había 18 programas educativos a nivel nacional y una matrícula de 1,082 inscritos y 31 egresados en México. Oficialmente, según la misma fuente, existían los siguientes programas específicos:

2 Doctorados en Inteligencia Artificial (1 pública, 1 privada) con 15 inscritos y 0 egresados; 2 Ingenierías en Inteligencia Artificial (2 públicas) con 498 inscritos y 0 egresados; 6 Licenciaturas en Inteligencia Artificial (2 públicas, 4 particulares) 221 inscritos y 10 egresados; 8 Maestrías en Inteligencia Artificial (5 públicas, 3 particulares) con 341 inscritos y 21 egresados; 1 TSU (técnico superior universitario) en Tecnologías de la Información en el área de Inteligencia Artificial con 7 inscritos.

México, incluso dentro de la región latinoamericana aparece con una oferta muy baja de programas de maestrías en IA en universidades de ranking QS, superándolo considerablemente Uruguay, Chile, y en menor medida, Argentina, Colombia y Ecuador (CEPAL, 2024).

Otra área que requiere ser abordada para que pueda operar plenamente la inteligencia artificial en México y que ahora padece grandes lagunas, es la política pública, junto con el marco legal, institucional y regulatorio que encuadre esta tecnología. Una iniciativa específica para este fin es la Alianza Nacional de Inteligencia Artificial (ANIA), que surgió en abril de 2023 como resultado de un acuerdo de trabajo entre la Comisión de Ciencia y Tecnología de la Cámara de Senadores y la Comisión de Ciencia, Tecnologías e innovación de la Cámara de Diputados en el Congreso Mexicano, en la que participan el sector público, el privado, la academia y la sociedad civil. El resultado de este esfuerzo –*Propuesta de Agenda Nacional de Inteligencia Artificial (2024-2030)*– se resume en el Recuadro 1 y da una visión de la enorme tarea que México necesita llevar a cabo para enfrentar el reto de la adopción y uso de la inteligencia artificial en el país.

RECUADRO 1 Propuesta de Agenda Nacional de Inteligencia Artificial (2024-2030) (mayo 2024). Síntesis (propia)

La propuesta de ANIA incluye recomendaciones de política pública; de regulación; de gobernanza; y de construcción de indicadores para la IA en México. **A. Las recomendaciones de Política Pública** incluyen (a) incorporación de la promoción de la IA a través de programas de política pública, como el Plan Nacional de Desarrollo (PND), la Estrategia Nacional Digital (EDN), y el Plan de Cumplimiento de las Metas de Desarrollo Sostenible 2030; (b) incorporar la IA como instrumento de educación y capacitación en mercados laborales, pero que, a su vez, capacite al capital humano en el uso de IA en sus actividades: implica el impulso de IA en los programas de educación básica, media superior y superior, así como educación continua y otros programas educativos, incluyendo la creación de habilidades de acuerdo a los requerimientos de la tecnología basada en IA; al mismo tiempo, se recomienda el incremento de I & D de tecnologías emergentes para impulsar el desarrollo; (c) fortalecer la ciberseguridad y la gestión de riesgos: propone la creación de una Ley de Ciberseguridad que incluya la capacitación y la adopción de protocolos para sistemas de protección contra ciberataques que protejan tanto al sector público como al privado; (d) aborda aspectos de género, inclusión y responsabilidad social en el uso de IA; (e) propone fortalecer la infraestructura y recolección de bases de datos: considera promover la inversión en el desarrollo de la infraestructura pública digital, incluyendo redes de telecomunicaciones, software, modelos de IA y estructuras de datos abiertos diseñadas bajo principios éticos (falta una política nacional de gobernanza de datos); (f) Innovación, investigación e industria: Buscar formas de atraer más capital a las actividades de I & D, incluyendo el campo de la IA, de forma que se alcance el 2.3% del PIB, promedio de la OCDE en I & D (pues México actualmente sólo llega al 0.3% del PIB. Contempla, entre otras cosas, el desarrollo de la investigación colaborativa a través de proyectos conjuntos entre universidades, industrias y emprendimientos de base tecnológica para avanzar en el desarrollo de la IA.
B. Recomendaciones de Regulación: Es muy amplia la regulación que se requiere para asegurar un uso adecuado de la IA en una sociedad, mucha de ella en proceso de creación o aún del todo ausente de la legislación y marco regulatorio mexicano. El ámbito de esta regulación debe abordar temas de derechos humanos; organismos certificadores de software y hardware para el desarrollo, uso y aplicación de las tecnologías basadas en IA; gestión de datos en la nube y centros de datos; regulación de la IA basada en la neutralidad tecnológica, entre otros principios; vincular las regulaciones y definiciones locales con las internacionales en materia de IA; contar con una Ley de Ciberseguridad sólida; fortalecer la protección de la propiedad intelectual en el ámbito de la operación de la IA; entre otras. **C. Recomendaciones de Gobernanza:** entre las recomendaciones más importantes de esta sección está la de integrar las atribuciones de la Coordinación de Estrategia Digital Nacional de la Presidencia de la República, y las del Centro de Investigación e Innovación en Tecnologías de la Información (INFOTEC) para crear la Agencia Digital Nacional (ADN) como organismo público descentralizado, dependiente del Ejecutivo Federal. **D. Recomendaciones de Indicadores:** se trata de medir en forma precisa los efectos positivos y negativos de la IA en México en áreas tales como economía, empleo, privacidad, y equidad. Para ello se precisa fortalecer capacidades institucionales para recabar datos confiables y generar indicadores adecuados. Entre otras cosas, se tiene la intención de generar una matriz nacional de indicadores de IA, consistente con los marcos de indicadores globales que permitan comparar el avance de México respecto a otros países en esta materia.

FUENTE: Lagunes, A. Martínez, Y., Cárdenas, C., De la Peña, S., Mancilla D., Xilotl, R., Sánchez, O., Moguel, A., Cárdenas, J. (2024), *Propuesta de Agenda Nacional de Inteligencia Artificial (2024-2030)*, disponible en ANIA, https://www.ania.org.mx/

Finalmente, aunque debería ser parte de las políticas públicas consideradas en la agenda de IA mencionada, la generación de energía para posibilitar el uso de la IA en la industria y específicamente en el sector manufacturero, merece especial atención. En la práctica, la disponibilidad de energía es insuficiente aún cuando el uso de la IA está en sus etapas iniciales en México y una introducción más generalizada de ella requeriría mayor disponibilidad de tal energía. A pesar de que, como ya se mencionó anteriormente, la IA ayuda a ahorrar este insumo a raíz de la mayor productividad que permite alcanzar en los procesos manufactureros, a la vez, requiere de un gran consumo de energía, indirectamente, al tener que procesar enormes cantidades de información en un creciente número de centros de datos, en los que se guarda información, se entrenan algoritmos y se aplican modelos de IA. El balance entre el ahorro y el consumo de energía por el uso de IA no está claro y difiere entre países y entre sectores productivos. Cuanto más avanzada la tecnología de IA, mayor es la cantidad de energía que requiere (esto es particularmente así para la IA generativa). Se estima que casi el 6% de la energía producida en Estados Unidos en 2026 se estará destinando a los centros de datos (en 2022 era el 4.5%) (WEF, 2024). Puede haber un ahorro significativo en el uso de energía por parte de la operación de IA, por ejemplo, usando chips ultra modernos, pero la capacidad de obtener y aplicar tecnología de última generación no está a la mano para todos los países ni para todos los sectores. Los requerimientos de electricidad por parte de la planta industrial en México ha estado creciendo a pasos agigantados debido a la multiplicación de los parque industriales y a la mayor necesidad de energía asociada al aumento de la producción (y, probablemente, el tipo de producción). De acuerdo a la Asociación de Parques Industriales Privados (AMPIP), en 2019 se tenían registrados 273 parques, mientras en 2022 había 432 y al mes de octubre de 2023, 460, y este ritmo había aumentado considerablemente en 2024 (especialmente en la franja norte del país) (SiiLa, 2023). La Norma Mexicana de Parques Industriales NMX-R-046-SCFI-2025 determina que el suministro mínimo debe ser de 150 kVAs por cada 10,000 m^2 [14], pero la misma fuente indica que el 38% de 144 plantas industriales evaluadas no tenían esa disponibilidad de energía. Las perspectivas de crecimiento era de 9% en plantas industriales entre 2023 y 2025, por lo que la necesidad de electricidad será aun mayor.

14 kVA: "cantidad total de Fuerza que un dispositivo consume. Se calcula multiplicando el voltaje (la intensidad de la Fuerza) por la corriente (la velocidad del flujo)." (Vogar, https://vogar.com.mx/blog/que-significan-los-kva-de-un-regulador)

El proyecto de aumento de energía eléctrica por parte del Programa de Desarrollo del Sistema Eléctrico Nacional (PRODECEN) era de 2.5% anuales entre 2023 y 2037, lo que es ya insuficiente para satisfacer la demanda de este sector en expansión. Un documento del IMCO que alude al programa más reciente –*Prodesen 2024-2038: El sistema eléctrico mexicano*– da indicios adicionales de las limitaciones en la generación de energía eléctrica en el país ya que señala que entre 2022 y 2023 la demanda eléctrica creció 3.5%, mientras que la capacidad de generación se incrementó únicamente 0.6%. El problema es particularmente agudo en las regiones con mayor crecimiento de la demanda energética y con poca infraestructura eléctrica como es el caso de la península de Yucatán.

El problema no es únicamente la generación de energía insuficiente sino también la calidad de esa energía en términos ambientales. El documento de IMCO mencionado apunta a que la generación limpia de energía está lejos de los propios objetivos de México en materia de transición energética, ya que su compromiso en el Acuerdo de París, e incluido en su legislación nacional, era alcanzar el 35% de energía generada a partir de fuentes limpias en 2024, pero este porcentaje sólo llegaba a 24.3% en 2023 (incluyendo generación distribuida). La meta de llegar al 43% de energía limpia en el total para el 2030 acordada por México se ve aún más difícil de cumplir. Nótese que este puede ser un desincentivo para las inversiones en el sector manufacturero, no sólo porque la energía parece escasear, sino también porque las compañías globales tienen compromisos de utilizar energía limpia para introducir sus productos a buena parte del mercado internacional. Si el uso de IA aumenta la presión en la demanda por energía, ésta también lo hará por energías limpias.

En la práctica, existen ejemplos de diversas empresas manufactureras operando en México que ya usan la IA en sus procesos. Entre éstas pueden mencionarse la empresa **Nemak**, un proveedor líder mundial de componentes de aluminio para la industria automotriz, que utiliza la IA especialmente para control de calidad, mantenimiento predictivo y optimización de procesos para mejorar la eficiencia y la calidad del producto. **Grupo Bimbo**, una de las empresas de panadería más grandes del mundo con operaciones en México, cuenta con IA en sus instalaciones de producción para optimizar las operaciones de la cadena de suministro y mejorar la eficiencia de la producción. La IA se utiliza para la previsión de la demanda, la gestión del inventario y la optimización de rutas. **Cemex**, una empresa global de materiales de construcción con sede en México, ha adoptado la IA para el mantenimiento predictivo en sus plantas de fabricación de cemento. Al analizar los datos de los sensores y equipos, Cemex

puede predecir las fallas de los equipos y programar el mantenimiento de manera proactiva para minimizar el tiempo de inactividad. **Grupo Modelo**, un productor de cerveza líder en México y parte de Anheuser-Busch InBev, utiliza la IA en sus procesos de elaboración de cerveza para optimizar la producción y el control de calidad. Se utilizan algoritmos de IA para supervisar los procesos de fermentación, analizar datos y realizar ajustes en tiempo real para la eficiencia de la elaboración de la cerveza.[15]

La industria automotriz es muy fuerte en México y sigue atrayendo nuevas inversiones. En las últimos dos años, a nivel mundial, este es el sector manufacturero en el cual más se ha acelerado la inversión en la IA, según un estudio reciente de IBM (IBM, 2023). Dentro de la industria manufacturera mexicana, la industria automotriz es la que, por ahora, más aplica la inteligencia artificial en su producción tanto para la optimización de la cadena de suministro, la planificación de producción, como en la detección temprana de fallas en vehículos. Una de las mayores fábricas automotrices en México, **Nissan** en Aguascalientes, utiliza la IA para optimizar las líneas de producción. Esta planta, una de las más grandes de Nissan en todo el mundo, emplea algoritmos de IA para gestionar robots de montaje, predecir las necesidades de mantenimiento y mejorar la eficiencia en las tareas de producción. Los automóviles mismos son cada vez más inteligentes: los vehículos pueden percibir, razonar y tomar decisiones, a través de la integración de sistemas electrónicos y algoritmos de aprendizaje automático. Ello permite al vehículo procesar grandes cantidades de datos y tomar decisiones en tiempo real.[16]

En una encuesta reciente (2024) a 87 líderes de la industria manufacturera en México (automotriz, industrial, electrónico, aeroespacial, productos médicos, farmacéutica, textil, bienes de consumo,etc.), realizada por KPMG (KPMG, 2024), al preguntárseles sobre el uso de IA en sus empresas, 35% consideró que será crucial su adopción a largo plazo, 29% a mediano plazo y 18% a corto plazo. Los propósitos de adopción, de mayor a menor importancia fueron: la optimización o reducción de costos: 73%; la ciberseguridad: 57%; gestión de la cadena de suministro: 55%; toma de decisiones basadas en datos: 55%; Mejorar la experiencia de clientes y proveedores: 45%. El tema de la IA, entonces, está empezando a penetrar en la industria manufacturera mexicana y ya está en las estrategias tecnológicas de las empresas, especialmente de las grandes.

15 Los ejemplos de empresas manufactureras que utilizan IA en México se obtuvieron mediante el uso del programa de GPT-4 de OpenIA. Consultado durante octubre de 2024.

16 Misma fuente, GPT-4, citada para otras empresas mexicanas que usan AI.

Es necesario considerar que probablemente la tecnología de punta que utiliza IA se irá adoptando a una mayor velocidad en México no sólo porque está penetrando en la manufactura mundial, como se describió en la primera parte de este capítulo, sino porque, además, dentro del contexto geopolítico internacional, el país está atrayendo inversiones en la industria de tecnología de frontera. Existe un esfuerzo importante de desacoplamiento entre Estados Unidos y China que se manifiesta, entre otras cosas, en una política comercial agresiva por parte del primer país para evitar ser desplazado por el segundo en la producción de la industria tecnológicamente más avanzada.

Más allá de las empresas manufactureras que ya tienen presencia importante en México, dentro de la gama de empresas que buscan reubicarse o expandir su producción en territorios amigables para Estados Unidos, destaca la industria de semiconductores, especialmente protegida por Estados Unidos para evitar la transferencia de tecnología de los chips de última generación a China (Schatan, 2024). Ya hay señales de interés en este sentido por parte de grandes empresas globales, como por ejemplo las inversiones planeadas por la empresa taiwanesa Foxconn (Forbes, 21/10/2024), anunciadas a principios de octubre de 2024, y que sería la mayor planta del mundo para ensamblar los superchips GB200 de Nvidia (empresa estadounidense), un componente clave de la plataforma informática de próxima generación (esencial para la operación de la IA) de la familia Blackwell. Esta empresa productora de los chips muy avanzados requeridos para operar la IA, a su vez necesita de la IA para el funcionamiento de sus empresas en México.

Si bien México ya produce semiconductores, su presencia es aún limitada en esta industria y, siendo un gran consumidor de estos dispositivos, por ser pieza clave en muchos de sus productos de exportación (automóviles, productos electrónicos, dispositivos médicos, etc.) tiene un déficit comercial importante en este rubro. Su potencial para convertirse en un importante productor de chips para surtir sus propias empresas manufactureras, así como un relevante exportador de chips sofisticados especialmente a su vecino del norte, aprovechando su ubicación geográfica, y algunas otras ventajas con las que cuenta, como mencionado, es prometedor. En la cadena de valor de semiconductores –diseño, fabricación y ensamblaje de componentes electrónicos– México tiene las mayores posibilidades de incorporarse a la última fase de la cadena de valor (aunque ya cuenta también con algunas actividades en diseño). Esta etapa –empaquetado y pruebas de semiconductores (OSAT, por sus siglas en inglés)– es el más intensivo en mano de obra en la cadena de producción, pero el uso de IA en ella es creciente y muy importante, por lo cual se necesita contar con la infraestructura necesaria para ello. De hecho,

los proveedores de OSAT están incorporando cada vez más tecnologías avanzadas en sus procesos de integración de semiconductores. Entre otras cosas, usan IA y aprendizaje de las máquinas para optimizar los procesos de prueba, mejorar la detección de defectos y afinar el control de calidad en la fabricación de semiconductores.[17]

En síntesis, el sector manufacturero de México comienza a estar familiarizado con la utilidad que les pudiera significar el uso de IA en muy diversas operaciones de sus empresas. Ya hay numerosos ejemplos de empresas, especialmente las globales, que utilizan en sus operaciones la IA en el país y hay perspectivas de importantes nuevas inversiones o la ampliación de las existentes en sectores que realmente requieren adoptar de la tecnología inteligente y que forman parte de la ola de inversiones que están llegando en el marco del *nearshoring*. Los desafíos para el país son grandes dado su rezago en varias capacidades que se necesitan para hacer posible el uso creciente y eficiente de la IA: la formación de capital humano, la I&D en este campo, la disponibilidad de energía eléctrica suficiente y progresivamente más limpia, el marco legal y regulatorio capaz de ordenar y garantizar el uso seguro de la IA para todas las partes involucradas, entre otras.

VII. REFLEXIONES FINALES

En este capítulo hemos hecho un análisis de cómo la introducción de la IA está revolucionando toda la cadena de valor de la producción manufacturera, especialmente de la Industria 4.0, y hasta qué punto puede transformar cada etapa de dicha cadena, desde aquella del diseño del producto hasta la distribución final de éste, pasando por el proceso mismo de fabricación, el manejo de inventarios, el mantenimiento de la maquinaria y el control de calidad del bien producido. En buena parte de este proceso, la IA permite alcanzar niveles de productividad sin precedentes y conseguir un grado de precisión del que los humanos por sí mismos no son capaces, o bien, requieren de muchos más recursos que si se apoyan en la IA. El uso de esta tecnología, sin embargo, trae aparejado una serie de retos muy grandes como el gran uso de energía que implica, las emisiones vinculadas a ella (dependiendo de qué tan limpias o sucias son las fuentes que la generan), además de requerir de capital

17 Medium (29/07/2024), *The Role of OSAT in Advancing Semiconductor Technology*, https://medium.com/@abhi.knp82/the-role-of-osat-in-advancing-semiconductor-technology-5c299847e733

humano, actualmente escaso, para operar esta nueva tecnología (aunque la propia IA proporciona herramientas de formación y capacitación que pueden ser extraordinariamente útiles en las empresas en que las aplican). A ello hay que agregar elementos, que no se trataron explícitamente en el trabajo, como la gran necesidad de inversiones que el introducir la IA requiere en maquinaria, equipo y capacidad computacional en las empresas (o en centros de datos en este último caso) y que no está al alcance de todas las empresas ni de todos los países, especialmente de aquellos en vías de desarrollo.

Se analizó específicamente las posibilidades de México de adoptar la IA en su sector manufacturero. Se argumentó que las posibilidades del *nearshoring* a raíz de las tensiones geopolíticas internacionales pueden favorecer la expansión de la industria 4.0, como el sector automotriz, el electrónico, y el de semiconductores específicamente, para lo cual se necesita estar mucho mejor preparados para la asimilación de la IA que lo que está el país actualmente. Aún cuando México se posiciona favorablemente en algunos indicadores para la adopción de IA en ALC (CEPAL, 2024), esta región está en general bastante atrasada respecto de lo que se necesita en infraestructura, capital humano, y varios otros elementos indispensables para el funcionamiento de la manufactura de última generación. Si bien la información estadística es escasa para obtener una idea precisa de la capacidad de asimilación de la IA por parte del sector manufacturero mexicano, varios indicadores sugieren que el país necesita tener una mejor preparación en diversos frentes: destacan la necesidad de una infraestructura de comunicación digital más desarrollada, incluyendo el contar con suficientes redes 5G de óptima calidad de manera que pueda funcionar la IA en la industria manufacturera que la requiera. También se necesita dar un impulso a la educación en todos los niveles, que prepare el capital humano para esta nueva era productiva –de las empresas inteligentes– en la cual México está incursionando, pero que aún está lejos de lograr. Asimismo, considerando la gran ampliación de las plantas industriales que se está percibiendo y las inversiones extranjeras que se están ya registrando o anunciando, los sectores manufactureros de tecnología de última generación requieren de una serie de insumos indispensables como energía (creciente y cada vez más limpia) y agua, además de un ambiente de seguridad jurídica y pública.

VIII. BIBLIOGRAFÍA

Abd El Hedi Gabsi (2024), "Integrating artificial intelligence in industry 4.0: insights, challenges, and future prospects–a literature review", *Annals of Operations Research*, disponible en: https://doi.org/10.1007/s10479-024-06012-6, Springer.

Berg, L. P., & Vance, J. M. (2017). "Industry use of virtual reality in product design and manufacturing: a survey". *Virtual Reality*, 21, 1-17, Springer.

Bousdekis, Alexandros, Katerina Lepenioti, Dimitris Apostolou, Gregoris Mentzas (2019), *Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0*, IFAC, Elsevier Ltd.

Comisión Económica para América Latina y el Caribe (CEPAL) (2024), *Indice Latinoamericano de Inteligencia Artificial* (ILIA), disponible en: https://indicelatam.cl

Deloitte (2023), *Managing a modern supply chain: Utilizing AI to combat complexity*, disponible en: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-managing-a-modern-supply-chain.pdf

Forbes (21/10/2024), *Además de Foxconn, 4 empresas construirán fábricas de chips en Jalisco*, disponible en: https://forbes.com.mx/ademas-de-foxconn-4-empresas-construiran-fabricas-de-chips-en-jalisco/

Huang, Chao; Siqi Bu; Hiu Hung Lee; Chun Hung Chan; Shu Wa Kong; Winco K.C. Yung (2024),

"Prognostics and health management for predictive maintenance: A review", *Journal of Manufacturing Systems*, Vol. 75, agosto, disponible en: https://www.sciencedirect.com/science/article/abs/pii/S0278612524001183

Ibarra Ernesto, De la Peña Sissi y Santoyo Cristina (2024); *Panorama de la Inteligencia Artificial en México: hacia una Estrategia Nacional*, disponible en: https://www.amcid.org/page/sandboxregulatoriomexico.

IBM (2023), IBM Global AI Adoption Index – *Enterprise Report*, Morning Consult, 8-23 noviembre.

Instituto Federal de Telecomunicaciones (IFT) (2024), *Plan de Licitación N° IFT-12, servicio de acceso inalámbrico*, disponible en: https://www.ift.org.mx/sites/default/files/industria/temasrelevantes/22186/documentos/12licitacionift-12vcp.pdf

Instituto Mexicano para la Competitividad (IMCO) (2024), *Prodesen 2024-2038: El sistema eléctrico mexicano,* 2 de junio, disponible en: https://imco.org.mx/prodesen-2024-2038-el-sistema-electrico-mexicano/

KPMG (2024), *El nearshoring tiene el potencial de impulsar a la industria de manufactura: KPMG*, disponible en: https://kpmg.com/mx/es/home/sala-de-prensa/press-releases/2024/09/cp-el-nearshoring-tiene-el-potencial-de-impulsar-a-la-industria-de-manufactura-pero-tambien-incrementara-la-competencia-kpmg.html

Lagunes, A. Martínez, Y., Cárdenas, C., De la Peña, S., Mancilla D., Xilotl, R., Sánchez, O., Moguel, A., Cárdenas, J. (2024), *Propuesta de Agenda Nacional de Inteligencia Artificial (2024-2030)*, disponible en: ANIA, https://www.ania.org.mx/

Medium (2021), *Boeing: Cuts 75% training time with VR & Increases accuracy by 33% using AR*, disponible en: https://infivr.medium.com/-a11a38d22f0d

Plathottam, Siby Jose, Arin Rzonca, Rishi Lakhnori, Chukwunwike O. Iloeje (2023), "A review of artificial intelligence applications in manufacturing operations", *Journal of Advanced Manufacturing and Processing*, 16 de mayo, disponible en: https://aiche.onlinelibrary.wiley.com/doi/full/10.1002/amp2.10159

Ruiz-Sarmiento, Jose-Raúl y Javier Gonzalez-Jimenez (2024), "MachNet, a general Deep Learning architecture for Predictive Maintenance within the industry 4.0 paradigm", *Engineering Applications of Artificial Intelligence,* Vol. 127, Parte B, enero.

Schatan, Claudia (2024), "Geopolítica, Políticas Industriales Digitales y la Importancia de la Industria de los Microchips", en Arturo Oropeza García, (coordinador) *Hacia una Visión de Desarrollo Industrial Digital en México*, INADI y CCE, disponible en: https://www.inadi.mx/_files/ugd/527291_b27deff64b5f47d3a57d2eb2fa4bd504.pdf

Schatan, Claudia (2023), "El despliegue de la banda 5G y su potencial para la industria 4.0", *Ciencia*, Vol. 74, N° 2, abril-junio, disponible en: https://www.revistaciencia.amc.edu.mx/index.php/vol-74-numero-2

SiiLa (2023), *Retos y Perspectivas en el suministro eléctrico del sector industrial mexicano; análisis de kVAs*, disponible en: https://siila.com.mx/noticias/retos-perspectivas-suministro-electrico-sector-industrial-mexicano-analisis-kvas/7092/lang/es

Sundaran, Sarvech y Abe Zeid (2023), "Artificial Intelligence-Based Smart Quality Inspection for Manufacturing", *Micromachines*, disponible en: https://www.mdpi.com/2072-666X/14/3/570

Tortoise Media, *The Global AI Index 2024*, disponible en: https://www.tortoisemedia.com/intelligence/global-ai/

World Economic Forum (WEF) (2024), *AI and energy: Will AI help reduce emissions or increase demand? Here's what to know*, 22 julio, disponible en: https://www.weforum.org/stories/2024/07/generative-ai-energy-emissions/

World Intellectual Property Organization (WIPO) (2024), Global Innovation Index, 2024, Mexico ranking in the Global Innovation Index 2024.

World Resources Institute (2019), *Los compromisos climáticos de México*, disponible en: https://es.wri.org/noticias/los-compromisos-climaticos-de-mexico#:~:text=México%20se%20comprometió%20a%20que,de%20carbono%20negro%2C%20requieren%20incentivos.

INTELIGENCIA ARTIFICIAL INDUSTRIAL EN CHINA

SUMARIO

I. Inteligencia artificial para la industrialización **II.** Concepto de Inteligencia Artificial Industrial **III.** La Inteligencia Artificial Industrial en China **IV.** Situación actual **V.** Perspectivas y conclusiones **VI.** Bibliografía.

> BEATRIZ JUÁREZ AGUILAR
Profesora de la División de Estudios de Posgrado de la Facultad de Derecho, UNAM. Consejera del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI)

I. INTELIGENCIA ARTIFICIAL PARA LA INDUSTRIALIZACIÓN

El 21 de enero de 2025 el gobierno estadounidense anunció la creación de *Stargate Project*, una empresa que construirá infraestructura para inteligencia artificial (IA) en los Estados Unidos. Este proyecto busca "asegurar el liderazgo estadounidense en inteligencia artificial", "apoyar la reindustrialización de Estados Unidos", además de proporcionar "capacidad estratégica para proteger la seguridad nacional" de este país y sus aliados. Sus socios tecnológicos clave inicialmente son Arm, Microsoft, NVIDIA, Oracle y OpenAI, quienes desarrollarán IA y en particular inteligencia artificial general, con una inversión planeada de 500 mil millones de dólares durante los próximos cuatro años, "en beneficio de toda la humanidad" (OpenAI & SoftBank, 2025).

El nuevo énfasis de los gobiernos de países desarrollados en la reindustrialización –explica Philip Pilkington– se debe a "los excesos" causados por la Nueva Economía de los años 1990, la economía postindustrial y las dos crisis financieras de 2000 y 2008, que junto con las tensiones geopolíticas actuales han conducido a la exploración de la reindustrialización bajo conceptos como la "'relocalización'" y el "'desacoplamiento'" (2024).

El empleo de la inteligencia artificial para apoyar la reindustrialización se debe a que "la reindustrialización puede asumir diferentes formas"; pudiendo ser la "relocalización de actividades que ya existían", "o de la creación de nuevas actividades industriales" (Capello & Cerisola, 2022).

En las recomendaciones a Occidente, Pilkington sugiere que deben dejarse de lado "soluciones rápidas" –especialmente las conducentes "a niveles peligrosos de conflicto global"– y enfocarse en "desarrollar un plan de varias décadas para impulsar masivamente inversión" (2024).

Esto es precisamente en lo que China ya está trabajando: un plan de modernización industrial integral a más de diez años, donde una herramienta esencial es la inteligencia artificial.

Este capítulo presentará un primer acercamiento.

DELIMITACIÓN

Este estudio analizará la política pública nacional china, ya que ésta indica los objetivos principales del gobierno y sirve de guía para autoridades nacionales, regionales y locales, así como para empresas nacionales e internacionales. Se revisará la política pública sobre la industria de 2022-2024 que tenga una relación específica tanto con la Nueva Industrialización como con la inteligencia artificial industrial. Pero dada la complejidad y la enorme cantidad de material, el análisis se centrará en la manufactura, la cual se considera el área central de la industria y de la economía.

Cabe señalar que la creciente confrontación geopolítica entre EUA y China, ha vuelto a la inteligencia artificial un tema muy delicado, por lo que China es muy selectiva en la información que publica.

Por cuestiones de espacio, este análisis tampoco abarcará la compleja relación entre los actores intraindustriales (por ejemplo, los fabricantes de equipo) o las relaciones entre la cadena de suministros.

II. CONCEPTO DE INTELIGENCIA ARTIFICIAL INDUSTRIAL

La Inteligencia Artificial Industrial (IAI) es una especialización de la inteligencia artificial que se aplica en entornos industriales (IFS, S/F); esta aplicación más enfocada no busca el desarrollo de sistemas con inteligencia similar a los humanos (Precognize, 2022). En sí, es una disciplina sistemática enfocada "en el desarrollo, la validación, implementación y mantenimiento de soluciones de inteligencia artificial (en sus múltiples formas) para aplicaciones industriales con un rendimiento continuo" (Peres y otros, 2020: 220122).

De hecho, esta última característica es central, porque la IAI "tiene tolerancia cero a los falsos positivos o negativos, las respuestas tardías o las predicciones que generen desconfianza" (Precognize, 2022). Entonces, tener una inteligencia artificial de grado industrial implica "un nivel de calidad" que sea confiable, seguro y que esté "diseñado para alcanzar requerimientos y estándares rigurosos de los medios profesionales más demandantes" (Siemens 2024).

De esta forma, la IAI contribuye a la construcción de sistemas industriales inteligentes y resilientes (Lee, Singh & Azamfar, 2019: 5). En particular, contribuye a incrementar la eficiencia ya que reduce los errores humanos, agiliza la producción y mejora la eficiencia operacional; impacta el mantenimiento predictivo, porque al analizar los datos sobre el equipo, predice fallas antes de que ocurran, reduciendo el tiempo de inactividad y previniendo reparaciones costosas; permite mayores opciones de personalización, al flexibilizar la producción y dar una respuesta más rápida a los cambios del mercado (IFS, S/F); también, facilita la automatización de flujos de trabajo, aumenta la seguridad, mejora la gestión de la cadena de suministro y el rendimiento, eliminando obstáculos para alcanzar la eficiencia operativa. Por todo lo cual, impulsa la rentabilidad, el mantenimiento de las ventajas competitivas a largo plazo (Precognize, 2022), así como la toma de decisiones, la productividad, y también contribuye a impulsar la innovación. De esta forma, la IAI se vuelve "una herramienta crucial para la transformación industrial" (IFS, S/F). Una contribución extra, es que mientras las aplicaciones de IAI optimizan el rendimiento de los sistemas industriales, también inspiran avances teóricos en el campo de la inteligencia artificial (Palhares, Yuan & Wang, 2019: 9636).

Es importante señalar que cada sector industrial puede tener usos diferentes de la Inteligencia Artificial Industrial. En el caso de la manufactura, la aplicación de inteligencia artificial puede dividirse en tres sectores: la capa básica superior, media e inferior. La primera incluye software y hardware industrial como en infraestructura y equipo industrial inteligente; la capa intermedia del sistema incluye sistemas de control industrial y plataformas de internet industrial; y la capa de aplicación puede aplicarse a comunicaciones electrónicas, energía y electricidad y se utiliza principalmente en diseño, operaciones in situ, previsión de ventas, conservación de energía, reducción de emisiones, etc. (36 Krypton Research Institute, 2024). A la integración de la manufactura con la nueva generación de tecnologías de la información se les llama: manufactura inteligente (Dù 2024).

Por ejemplo, en la automatización se parte de que una máquina, un equipo o un robot realiza cierta actividad predefinida, siguiendo reglas específicas en un número limitado de escenarios. El contar con inteligencia artificial posibilita el reconocimiento de patrones dentro de los datos, se aprende de la experiencia y se mejora el rendimiento futuro, por lo que la IA permite que estas máquinas, equipos o robots, identifiquen soluciones a problemas complejos dentro de un espacio determinado (Lee, Singh & Azamfar, 2019: 5).

REQUISITOS GENERALES

La inteligencia artificial industrial precisa de un conjunto de tecnologías avanzadas para su desempeño efectivo, entre las cuales destacan:

A. el aprendizaje automático y aprendizaje profundo: los cuales permiten que "los sistemas aprendan de los datos, mejoren con el tiempo y hagan predicciones informadas, impulsando la eficiencia en los procesos industriales".

B. El internet de las cosas (IoT): los dispositivos "recopilan grandes cantidades de datos de máquinas y sistemas, que la IA analiza para optimizar los procesos, reducir el desperdicio y mejorar la productividad".

C. Robótica y automatización: "la robótica está en el corazón de la IA industrial", puesto que "permite la automatización de tareas que van desde el trabajo en la línea de montaje hasta las inspecciones de calidad".

D. Cómputo al borde (*Edge computing*): al procesar datos de forma local, "permite a los sistemas de inteligencia artificial tomar decisiones más rápidas sin necesidad de depender de la computación en la nube", lo cual es fundamental para las tareas industriales donde el tiempo es un factor central. (IFS, S/F).

A pesar del gran potencial de la Inteligencia Artificial Industrial, su implementación tiene muchos retos. Una primera razón básica es la gran inversión necesaria para realizar las adecuaciones para la integración de las aplicaciones de IA en las estructuras corporativas y a lo largo de toda la cadena de valor. Por lo que se recomienda a las compañías, no solo comprender su potencial, sino también los requisitos principales y los cambios organizativos necesarios para aprovechar todo su potencial (Peres y otros, 2020: 220122-220124).

Industrial and Financial Systems (IFS), empresa sueca de software empresarial, destaca entre las tareas centrales necesarias para la implementación exitosa y provechosa de la IAI a las siguientes:

A. Recolección y procesamiento de datos: debido a que la IAI precisa grandes cantidades de datos, para muchas empresas representa un reto la recolección, organización y análisis efectivo de estos datos.

B. Integración de sistemas heredados (*legacy systems*): la existencia de sistemas industriales antiguos que sean incompatibles con la IAI puede volverse un obstáculo.

C. Competencias de los trabajadores: los trabajadores, precisan nuevas habilidades y sus empleadores precisan invertir en el reentrenamiento.

D. Seguridad y privacidad: ante amenazas cibernéticas, la protección de los datos recolectados y por analizar se convierte en una prioridad esencial (IFS, S/F).

En este sentido, la investigación de Peres y otros, subraya que un obstáculo crítico en la transformación digital y la adopción tecnológica, es que muchas empresas desconocen la forma para integrar a la IAI a sus modelos y procesos de negocio, por lo que se precisa una guía de cómo hacerlo, con prácticas comunes, evidencia, experiencias exitosas. Por ejemplo, las empresas suelen recolectar datos en múltiples formatos, sin estándares y con diferentes grados de calidad; por lo que se precisan acercamientos sistemáticos para la implementación, despliegue y gestión de soluciones de IAI (2020: 220122-220124).

Un instrumento útil para solventar este problema son las compañías-faro, las cuales tienen como activo principal a la inteligencia, misma que se aprovecha plenamente mediante la adaptación y aplicación de la IAI, volviéndola un motor principal de desarrollo. La incorporación de tecnologías avanzadas en estas compañías, les ha traído beneficios económicos significativos, volviéndose un "verdadero faro" (Lee, Singh & Azamfar, 2019: 8-9) que ilumina y muestra el camino a otros.

III. LA INTELIGENCIA ARTIFICIAL INDUSTRIAL EN CHINA

A. PANORAMA GENERAL DE LA GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL

Dentro de la esfera económica china, la gobernanza de la inteligencia artificial contempla las mismas características que aquella que rige al resto del internet. En un primer momento, ambas se caracterizan por constar de múltiples partes interesadas (*multi-stakeholder*) centradas en un gobierno fuerte, donde la industria, la academia y la sociedad tienen intereses propios que hacen valer a través de negociaciones, demandas y luchas. Operan dentro de un ecosistema tecnológico consolidado y con innovación tecnológica, donde las grandes empresas tecnológicas nacionales y extranjeras acatan las reglas. Las políticas públicas son fuertes y determinan el rumbo a seguir, mientras que los instrumentos jurídicos y los estándares se adaptan a las primeras y trabajan conjuntamente en alcanzar los objetivos nacionales. El marco jurídico chino no solo permite la innovación tecnológica, sino que la impulsa (Juárez-Aguilar, en prensa, a).

En el caso particular de la inteligencia artificial, desde 2015, el gobierno impulsó varias políticas públicas tecno-económicas que señalaron el rumbo principal de la investigación, la innovación, el desarrollo y los escenarios de

aplicación de la IA. Meses más tarde, se expidieron una serie de marcos éticos voluntarios y de autorregulación industrial. Mientras tanto, empresas, gobierno y sociedad, adoptaron velozmente la IA por medio del comercio electrónico, la economía colaborativa, los juegos, fintech, redes sociales, entre muchos otros. No obstante, el uso indiscriminado de la IA, comenzó a crear múltiples problemas sociales que la regulación existente y las medidas voluntarias (como los marcos éticos y la autorregulación) no pudieron solucionar. Por esta razón, desde 2022 el gobierno fue elaborando y aplicando paulatinamente normas específicas, centradas en atender los problemas más apremiantes de los actores involucrados, como los consumidores, los menores de edad, los adultos mayores, los trabajadores de la economía colaborativa, entre otros. (Juárez-Aguilar, en prensa, b).

Al mismo tiempo, EUA fue incrementando las restricciones al sector tecnológico chino, como la exportación de productos, servicios y datos e incluso a la movilidad laboral y a la inversión, centrándose específicamente en detener los avances chinos en inteligencia artificial. En respuesta, el gobierno chino enfatizó la relación entre seguridad y autosuficiencia tecnológica, al tiempo que reorientó la gobernanza de la tecnología a objetivos específicos como la fortaleza industrial y la seguridad nacional, desincentivando áreas más relacionadas con el ocio, como los juegos, el comercio electrónico, etc. (Creemers y otros, 2023).

Es en este punto, donde surgió el concepto de Nueva Industrialización que tiene un renovado énfasis con la inteligencia artificial.

B. NUEVA INDUSTRIALIZACIÓN + INTELIGENCIA ARTIFICIAL

Desde hace casi veinte años, el gobierno chino ha impulsado múltiples políticas públicas que trabajan para la integración de nuevas tecnologías con la economía real para impulsar el desarrollo y la innovación. No obstante, en años recientes ha resaltado la primacía de la economía real para el país, en particular, de la industria manufacturera.

Desde 2022, el gobierno chino ha ido formando el concepto de Nueva Industrialización (新型工业化, *xīnxíng gōngyèhuà*)1, el cual hace referencia a que las nuevas generaciones tecnológicas deben contribuir a modernizar todo el

1 A lo largo del documento, los términos en chino se acompañarán con su respectiva transliteración, siguiendo el estilo de la República Popular China: el *pīnyīn*. Para posibilitar una mejor visualización, los caracteres chinos no se insertarán en itálicas.

panorama industrial a fin de crear nuevos modelos de producción, de negocio y nuevas formas de crecimiento económico, anteponiendo la seguridad a los resultados económicos. Como meta para hacer realidad estos objetivos, se ha señalado al año 2035 (Juárez-Aguilar, 2024). Esta Nueva Industrialización reconoce la centralidad de la manufactura y del mercado interno y enlista una serie de medidas para la transformación industrial, entre las que destacan actualizar las industrias tradicionales; consolidar y ampliar industrias donde ya existan ventajas; impulsar y expandir industrias emergentes; diseñar industrias futuras de forma prospectiva; así como ahondar en la cooperación internacional en la materia (OICE, 2023).

Desde la academia china, varios expertos han coincidido en la importancia de incorporar de manera integral a la inteligencia artificial con la industria, a fin de alcanzar las metas de la Nueva Industrialización.

Wén Jūnléi (文钧雷), vicepresidente y profesor del Instituto de Investigación de Realidad Virtual, del Centro de Computación en la Nube de la Academia China de Ciencias, asevera que dado que la inteligencia artificial y la manufactura inteligente están al centro de la nueva competencia industrial y que la manufactura es el fundamento del país, su conjunción puede lograr que China se vuelva un poder manufacturero capaz de superar retos domésticos e internacionales, además de proveer un gran impulso para el desarrollo de nuevos campos económicos (Diario de Chengdu, 2022).

Dù Chuánzhōng (杜传忠), investigador especial en el Instituto de Investigación de Estrategias sobre el Desarrollo de Inteligencia Artificial de Nueva Generación de China y profesor y director del Instituto de Economía Industrial de la Universidad de Nankai, detalla cómo la inteligencia artificial habilita la nueva industrialización. Parte de su argumento revela que desde la "perspectiva de la permeabilidad, la inteligencia artificial, como tecnología de propósito general, es capaz de integrarse en todos los aspectos de la industria manufacturera y de las cadenas industriales ascendentes y descendentes"; desde la sinergia, la IA "conecta orgánicamente los campos de consumo, producción, circulación y distribución, lo cual ayuda a mejorar la coordinación, el orden y la eficiencia de todo el proceso de reproducción social"; y desde la innovación, "la IA puede lograr la auto-evolución y la autoactualización de la tecnología a través del aprendizaje profundo y la iteración rápida, lo que a su vez conduce a la optimización y actualización de toda la industria" (Dù, 2024).

Zhōu Jì (周济), profesor y expresidente de la Academia China de Ingeniería, enfatiza que la manufactura inteligente precisa tanto de 5G como del internet industrial para lograr la transformación digital, el desarrollo y la mejora

inteligente de la industria manufacturera, así como la innovación del modelo industrial. Por lo que "5G + internet industrial + inteligencia artificial" empoderarán la Nueva Industrialización, posibilitando la innovación en productos, en tecnología de producción, en el modelo industrial y en la integración de sistemas de manufactura, alcanzando así el clímax de la Cuarta Revolución Industrial (Chén, 2024).

POLÍTICAS PÚBLICAS

Los planes que se presentarán a continuación, se ordenan bajo el Plan Quinquenal vigente que es el "14° Plan Quinquenal para el Desarrollo Económico y Social Nacional de la República Popular China y los objetivos a largo plazo hasta el año 2035". Coinciden en que buscan aprovechar las oportunidades brindadas por la nueva revolución científica y tecnológica para apoyar el fortalecimiento industrial y alcanzar los objetivos de la Nueva Industrialización, pero cada uno se acerca desde diferentes ángulos.

El 30 de diciembre de 2023, el Ministerio de Industria y Tecnologías de la Información (MIIT, por sus siglas en inglés) y ocho Instituciones más, presentaron "*Opiniones guía para acelerar la transformación y actualización de las industrias manufactureras tradicionales*".[2] Este plan considera que la adopción y actualización activa permitirá mejorar "la resiliencia y seguridad de la cadena industrial y de suministro", con lo cual China no solo liderará en la ciencia y tecnología, sino en la transformación industrial, fortaleciendo su competitividad y posición en la división global de trabajo industrial. Se propone que para 2027, habrá una mejora en la industria manufacturera tradicional china, con un desarrollo de alto nivel, inteligente, ecológico e integrado. Entre las medidas a adoptar, se considera promover la transformación inteligente de las empresas y de las redes digitales, tomando en cuenta las características y necesidades particulares de diferentes industrias, acelerando la integración de todo el proceso y elementos de la manufactura con la inteligencia artificial, big data, computación en la nube, 5G, internet de las cosas, entre otros (CAICT, 2023).

El 31 de enero de 2024, el MITT y otras seis entidades presentaron *Opiniones sobre la implementación del impulso al desarrollo e innovación en las industrias del futuro,*[3] el cual busca apuntalar a la manufactura desde el desarrollo y

2 《关于加快传统制造业转型升级的指导意见》*guānyú jiākuài chuántǒng zhìzào yè zhuǎnxíng shēngjí de zhǐdǎo yìjiàn.*

3 《关于推动未来产业创新发展的实施意见》*guānyú tuīdòng wèilái chǎnyè chuàngxīn fāzhǎn de shíshī yìjiàn*

aplicación de tecnologías de vanguardia y tecnologías disruptivas, por ejemplo el metaverso, los robots humanoides, las interfaces cerebro-computadora y la información cuántica. Especialmente sobre "la manufactura futura" busca desarrollar la manufactura inteligente, la biológica, la nano manufactura, la láser, la circular; así como tener avances disruptivos en la manufactura inteligente, la detección inteligente, la simulación analógica y otras tecnologías clave; promover la manufactura flexible, la compartida y otros modelos, así como el desarrollo del internet industrial, el metaverso industrial, entre otros. Con objetivos para 2025 y 2027, asegura que "China tiene ventajas integrales, como un sistema industrial completo, una gran escala industrial y ricos escenarios de aplicación, lo que proporciona un terreno fértil para el futuro desarrollo industrial" (MITT + 6, 2024).

También en enero de 2024, circuló la noticia de que el Consejo de Estado de la RPC se reunió para trabajar en el ***Plan de Acción para impulsar la inteligencia artificial para potenciar la Nueva Industrialización,***[4] el cual a la fecha de redacción de este texto aún no ha sido publicado (enero 2025).

En dicha reunión se señaló la necesidad de coordinar el desarrollo de alta calidad con la seguridad de alto nivel; teniendo como el objetivo central a la manufactura inteligente; como medio principal, a la profunda integración de la inteligencia artificial con la manufactura; y como el motor, a los escenarios de aplicación. Entre otros objetivos clave destacan la necesidad de la actualización inteligente de las industrias clave, el desarrollo de productos inteligentes y el fortalecimiento de los sistemas de manufactura a alto nivel (Chinese Government Website, 2024).

El 11 de mayo de 2024, el Consejo de Estado aprobó el ***Plan de acción para la digitalización de la industria manufacturera***[5], el cual enfatiza que su transformación no sólo implica una optimización industrial y del modelo de desarrollo, sino que está directamente relacionada con la seguridad económica nacional, la industrial, la científica y la tecnológica. Se acordó explorar escenarios típicos en diferentes campos, de acuerdo con las necesidades específicas de la industria. De la misma forma se precisa apresurar la investigación, promoción y aplicación de avances en las tecnologías clave; trabajar en la conexión de equipos, el reconocimiento mutuo de protocolos, la formulación de estándares, la construcción de plataformas, entre otros. Asimismo se recalcó la importancia

4 《推动人工智能赋能新型工业化行动方案》 *tuīdòng réngōng zhìnéng fù néng xīnxíng gōngyèhuà xíngdòng fāng'àn*

5 《制造业数字化转型行动方案》 *zhìzào yè shùzìhuà zhuǎnxíng xíngdòng fāng'àn*

de incrementar los mecanismos de apoyo a largo plazo para la transformación digital de las pequeñas y medianas empresas y combinarlos con acciones a gran escala de renovación de equipos y la implementación de proyectos de transformación y modernización tecnológica, así como mejorar las plataformas de servicios públicos (Xinhua, 2024).

En noviembre de 2024, Hé Hailín (何海林), jefe del Buró de Monitoreo de Operaciones y Coordinación del MIIT, aseguró que se está trabajando en el fortalecimiento y la coordinación de políticas públicas, profundizado en la cooperación internacional y luchando para "contribuir a la fuerza china para construir un sistema industrial y cadenas de suministro globales que sean seguros, estables, flexibles, eficientes, abiertos, incluyentes y mutuamente benéficos" (Consejo de Estado, 2024).

Los pasos que se siguen son:

1. Guiar y promover activamente a las empresas para que se integren en las cadenas industriales y de suministro internacional.
2. Fortalecer la reputación de la marca *Hecho en China* a través de múltiples medidas como el empuje que da el diseño industrial al mejoramiento de la calidad.
3. Desarrollar vigorosamente el comercio digital y el comercio verde. En cuanto al primero, se debe trabajar en los planes de acción diseñados para la transformación digital de la industria manufacturera y para que la inteligencia artificial potencie la Nueva Industrialización. Construir fábricas y cadenas de suministro inteligentes, desarrollar equipos logísticos inteligentes, promover la construcción de normas de comercio electrónico que sean compatibles con normas internacionales sobre flujo de datos, trazabilidad, firmas electrónicas, así como ampliar nuevos espacios para el comercio digital.
4. Mejorar el nivel de servicio de las empresas que "salen al extranjero" (Consejo de Estado, 2024).

IV. SITUACIÓN ACTUAL

En noviembre de 2024, el gobierno chino presentó el *Reporte de Desarrollo del Internet en China*,[6] el cual especificó que este país cuenta con la red 5G más grande del mundo que ha integrado a 74 categorías de la economía nacional, acelerando en áreas clave como la manufactura, minería, energía eléctrica, atención médica, entre otras. El número de empresas centrales de IA son más de 4500 y se han construido casi 10 mil talleres digitales y fábricas inteligentes; 421 de estas últimas sirven como compañías-faro y más del 90% cuentan con IA y gemelos digitales (CCTIME, 2024; Qinghai Daily, 2024).

De acuerdo con la International Federation of Robotics, China es el mercado más grande de robots industriales con 276,288 unidades instaladas en 2023, lo que representa el 51% global. El potencial de crecimiento en el sector de manufactura prevé un crecimiento anual entre el 5% y el 10% hasta 2027 (IFRI, 2024 B). No obstante, si se le ve desde la densidad de robots, en 2023 ocupó el tercer puesto con 470 robots por cada 10 mil empleados, solo detrás de Corea del Sur y Singapur. Sin embargo, este logro es sorprendente si se considera que China entró al top 10 hasta 2019, lo que implica que ha duplicado su densidad cada cuatro años (IFRI, 2024 A).

El *Instituto de Investigación 36 Krypton* (36氪研究院), presentó en mayo de 2024 el *Reporte de investigación de la industria manufacturera AI+ de China*,[7] el cual señala que la inteligencia artificial está transformando a la industria manufacturera de ese país porque busca satisfacer las necesidades personalizadas del usuario. Al enfocarse en el usuario, la producción se vuelve más flexible, dejando atrás la rigidez que implica centrarse principalmente en el producto (2024).

En consecuencia –asegura Dù Chuánzhōng–, la transformación digital de muchas empresas chinas de manufactura ha alcanzado un nivel tal que, la mera digitalización de las industrias tradicionales no es suficiente; sólo la integración profunda de la inteligencia artificial a la manufactura puede hacer realidad los objetivos de la Nueva Industrialización (2024).

6 《中国互联网发展报告2024》 *zhōngguó hùliánwǎng fāzhǎn bàogào 2024*

7 《中国AI+制造产业研究报告》 *zhōngguó AI+zhìzào chǎnyè yán jiù bàogào gǔn*

RETOS EN LA IMPLEMENTACIÓN DE LA INTELIGENCIA ARTIFICIAL INDUSTRIAL

Como se explicó en la segunda sección, la Inteligencia Artificial Industrial precisa de la recolección y procesamiento de datos, la integración de sistemas heredados, las adecuadas competencias de los trabajadores, así como mantener la seguridad y la privacidad (IFS, S/F).

Debido a que estas áreas son transversales e interactúan entre sí, muchas medidas y planes para su actualización y protección están también entrelazadas. A continuación un breve panorama.

A. Datos

En 2020, el gobierno chino designó a los datos como factor de producción, por lo que comenzó a crear las vías para que éstos apuntalaran la economía, por ejemplo los mercados de datos. No obstante, las confrontaciones geopolíticas han sido un elemento esencial para subordinar el desarrollo económico a la seguridad. De esta forma, desde 2023 se están estipulando instrumentos para regular las actividades de tratamiento de datos en el ámbito de la industria y las tecnologías de la información. Particularmente en el caso de los datos industriales, en primer lugar se deben determinar qué datos deben protegerse, para después indicar cuáles pueden utilizarse para promover el desarrollo económico, resguardando los derechos e intereses de individuos, organizaciones y la nación (Juárez-Aguilar, en prensa, c).

B. Integración con sistemas heredados (*legacy systems*)

Como se menciona en *Opiniones guía para acelerar la transformación y actualización de las industrias manufactureras tradicionales* (30/12/2023), las autoridades reconocen la necesidad de una actualización integral que incluya la infraestructura digital y el equipo de toda la industria, de las zonas de alta tecnología, de los parques industriales, etc. Por ejemplo, en cuanto a los equipos de producción se promueve "la aplicación de nuevos sensores, controles avanzados y otros componentes inteligentes", además de "acelerar la promoción de equipos inteligentes y la actualización y sustitución de software" (CAICT, 2023).

C. Habilidades de trabajadores

Si bien por décadas, China ha trabajado afanosamente en la adecuación de su sistema educativo para afrontar los retos científicos y tecnológicos que presenta la modernización, en la actualidad el Ministerio de Educación continúa con esa

labor y suele ser una institución convocante en muchas de las políticas públicas relacionadas con la transformación digital.[8]

En julio de 2024, esta institución publicó un documento donde se alentaba a las universidades a apoyar en áreas clave como "la fabricación de circuitos integrados, inteligencia artificial, tecnología cuántica, ciencias de la vida y energía", y se solicitó una mayor adecuación de la educación profesional al desarrollo de la industria. (Bangkok Post 2024). Entre las tareas señaladas por el Vice Ministerio de Educación, destacan: un ajuste dinámico de la profesión, diseño científico del contenido de los cursos, actualización de material didáctico, formulación de estándares para pasantías, construir una base de entrenamiento para integrar la industria y la educación, entre otros (CAEA, 2024).

En agosto de 2024, 19 universidades ya estaban en proceso de renovar "sus carreras de ingeniería y tecnología a favor de áreas de alta tecnología como IA y big data". De la misma forma, también se cuentan con programas de capacitación en big data, Inteligencia artificial, fabricación inteligente, circuitos integrados, seguridad de datos, entre otros, dirigidos a ingenieros, técnicos y trabajadores (Bangkok Post, 2024).

D. Seguridad

Ante los crecientes retos nacionales e internacionales que está enfrentando China, el gobierno ha decidido reforzar su seguridad de forma multidimensional: entre las áreas incluidas destacan la seguridad política, militar, territorial, económica, financiera, cultural, social, alimentaria, ecológica, así como la seguridad tecnológica que contiene a la inteligencia artificial, los datos, la infraestructura digital, entre otras (Juárez-Aguilar, en prensa, c).

Esta postura se ve reflejada en numerosas políticas públicas y un creciente marco jurídico; un ejemplo de las primeras es el mencionado *Plan de acción para la digitalización de la industria manufacturera* (2024) que entrelaza la optimización industrial con la seguridad económica nacional; y como ejemplo del marco jurídico, están las *Medidas para la gestión de la seguridad de los datos en el ámbito de la industria y las tecnologías de la información (para su aplicación experimental)*[9], el cual entró en vigencia el primero de enero de 2023.

8 Por ejemplo, en *Opiniones guía para acelerar la transformación y actualización de las industrias manufactureras tradicionales* de 2023, entre las 9 instituciones convocantes, una fue el Ministerio de Educación (CAICT, 2023).

9 工业和信息化领域数据安全管理办法（试行）*Gōngyè hé xìnxī huà lǐngyù shùjù ānquán guǎnlǐ bànfǎ (shìxíng)*

PERSPECTIVAS

Las transformaciones derivadas de las interacciones de la inteligencia artificial con su entorno, están dando lugar al surgimiento de nuevos retos y oportunidades. Por una parte, están apareciendo todo un grupo de escenarios de aplicación con riesgos a la seguridad, a la vez que posibles vulnerabilidades en infraestructura, software, hardware, privacidad, transmisión de datos, entre otros. Así que "comprender y gestionar los nuevos riesgos a la seguridad" tiene "una importancia estratégica cada vez mayor" (CCTIME, 2024).

Y no hay tiempo que perder, porque Jīn Zhuànglóng del MIIT ha señalado que esta institución ya apunta a nuevas direcciones: la fabricación futura, información futura y materiales futuros; robots humanoides, interfaces cerebro-computadora, metaverso, internet de próxima generación, 6G. además de que se buscan avances innovadores en tecnologías centrales clave, crear una serie de productos esenciales, obtener logros transcendentales, construir una serie de incubadoras de empresas, entre otros (Diario Guāngmíng, 2024).

V. CONCLUSIONES

En China, los cambios geopolíticos han alentado que el desarrollo se una profundamente a la seguridad y sus vertientes como la seguridad económica, la industrial, la científica y la tecnológica.

En este camino, este estudio ha mostrado que la modernización industrial china tiene una nueva vertiente donde se apoya de la inteligencia artificial; este tipo de fortalecimiento industrial se orienta tanto a actualizar la industria tradicional como a crear nuevas actividades industriales. Su abanico es muy amplio porque abarca desde la actualización, hasta los escenarios de aplicación y las industrias futuras; desde las empresas líderes mundiales hasta las PyMES; desde las industrias tradicionales hasta el desarrollo verde (planes de finales 2024).

Estas políticas públicas parten del concepto amplio de inteligencia artificial, que abarca desde infraestructura a los escenarios de aplicación, lo cual hace un acercamiento mucho más completo. Es palpable que estos instrumentos fueron hechos por expertos en Inteligencia Artificial Industrial, ya que incluyen todas sus áreas: datos, integración de sistemas heredados, competencias laborales, seguridad y privacidad.

China está posicionándose indudablemente en un nivel de transformación industrial puntero. Como señala el Prof. Dù Chuánzhōng, la industria china está alcanzando un nivel de digitalización tal, que sólo la integración profunda con inteligencia artificial puede cumplir los objetivos de la Nueva Industrialización (2024).

Dada la complejidad de cada industria y la heterogeneidad del panorama chino al nivel local, se pueden encontrar innumerables ejemplos. Y tal como el de las empresas faro, China nos muestra un camino posible para avanzar.

Un primer paso –indicado por el Prof. Zhōu Jì– es contar con 5G e internet industrial, a fin de aprovechar lo que big data, la nueva generación de inteligencia artificial y la manufactura inteligente puedan ofrecer (Chén, 2024).

Siempre y cuando tengamos en cuenta lo que subraya Philip Pilkington; cada industrialización sigue un camino diferente (2024).

VI. BIBLIOGRAFÍA

36 Krypton Research Institute 36氪研究院 (24/05/2024). Reporte de investigación de la industria manufacturera, AI+ de China, 中国AI+制造产业研究报告. Disponible en https://finance.sina.com.cn/tech/roll/2024-05-30/doc-inawzavy4636958.shtml

Bangkok post (23/08/2024). Universities in China change course to meet government's need for hi-tech workforce https://www.bangkokpost.com/world/2852553/universities-in-china-change-course-to-meet-governments-need-for-hi-tech-workforce

CAEA China Adult Education Association 中国成人教育协会 (25/11/2024). Wu Yan, Vice Minister of the Ministry of Education: Promote the high-quality development of vocational education through the integration of industry and education 教育部副部长吴岩：以产教融合推动职业教育高质量发展. https://www.caea.org.cn/newsinfo/7786781.html

CAICT (30/12/2023). *Opiniones guía para acelerar la transformación y actualización de las industrias manufactureras tradicionales*, 关于加快传统制造业转型升级的指导意见. https://mp.weixin.qq.com/s/CVjsdPHa5RkggX-nZRRAqQ

Capello, R., & Cerisola, S. (2022). Regional reindustrialization patterns and productivity growth in Europe. *Regional Studies*, 57(1), 1–12. https://doi.org/10.1080/00343404.2022.2050894

CCTIME (28/11/2024). Quick look! Here comes the focus of the Internet Development Report, 速看！互联网发展报告重点来了. https://finance.sina.com.cn/tech/roll/2024-11-28/doc-incxqqsy9863893.shtml

Chén Yǒngquán陈永权 (20/11/2024). Without a powerful 5G + industrial Internet, there will be no new generation of artificial intelligence (没有强大的5G+工业互联网，就没有新一代人工智能). Yangtze River Daily长江日报. http://m.ce.cn/cy/gd/202411/20/t20241120_39208886.shtml

Chinese Government Website 中国政府网 (27/01/2024). Relationship with real estate, RRR reduction, and artificial intelligence to empower new industrialization..... State Department Policy of the Week, 关系房地产、降准、人工智能赋能新型工业化……国务院本周政策https://www.gov.cn/yaowen/liebiao/202401/content_6928620.htm

Consejo de Estado RPC (22/11/2024). State Council policy briefing to respond to concerns 国务院政策例行吹风会, 回应关切https://www.gov.cn/xinwen/2024zccfh/27/index.htm

Creemers, Rogier, Johanna Costigan, Paul Triolo, Tom Nunlist, Lauren Dudley, Mei Danowski, Martin Chorzempa, Karman Lucero, Seaton Huang (2023). "Is China's Tech 'Crackdown' or 'Rectification' Over?". *DigiChina*. Disponible en https://digichina.stanford.edu/work/is-chinas-tech-crackdown-or-rectification-over/

Diario de Chengdu成都日报 (04/11/2022). How can AI empower manufacturing? (人工智能如何赋能制造业？). https://cdst.chengdu.gov.cn/cdkxjsj/c108732/2022-11/04/content_bcf3c911acb24e8ebd6c443e66870e64.shtml

Dù Chuánzhōng杜传忠(03/2024). Artificial intelligence fully empowers new industrialization (人工智能全面赋能新型工业化). *China Social Sciences Journal*. https://www.xuexi.cn/lgpage/detail/index.html?id=1737982780920682307&item_id=1737982780920682307

IFR International Federation of Robotics (20/11/2024 B). Global Robot Density in Factories Doubled in Seven Years. Disponible en https://ifr.org/ifr-press-releases/news/global-robot-density-in-factories-doubled-in-seven-years

IFR International Federation of Robotics (24/09/2024 A). Record 4 Million Robots on Factory Floors Worldwide. Disponible en https://ifr.org/downloads/press2018/2024-SEP-24_IFR_press_release_World_Robotics_2024_-_global_market.pdf

Industrial and Financial Systems IFS (S/F). What is Industrial AI™? A Comprehensive Overview. https://www.ifs.com/nl/what-is/what-is-industrial-ai Consultado (22/11/2024)

Jīn Zhuànglóng 金壮龙, (16/11/2024). Jīn Zhuànglóng: Further deepen the reform in the field of industry and information technology in an all-round way and inject strong impetus into the promotion of new industrialization (金壮龙：进一步全面深化工业和信息化领域改革　　为推进新型工业化注入强大动力). *Qiushi Magazine*《求是》杂志. https://www.gov.cn/zhengce/202411/content_6987599.htm

Juárez-Aguilar, Beatriz (2024). Después de Hecho en China 2025: la Nueva industrialización en China. En Arturo Oropeza García (Coord.), *Hacia una nueva visión del desarrollo industrial digital en México*. 2024. INADI-CCE, pp. 344-358. Disponible en https://www.inadi.mx/libros-inadi

Juárez-Aguilar, Beatriz (en prensa, a). El éxito de la gobernanza de la economía digital china.

Juárez-Aguilar, Beatriz (en prensa, b). La gobernanza de la inteligencia artificial en China.

Juárez-Aguilar, Beatriz (en prensa, c). Los datos industriales en China.

Lee, Jay, Jaskaran Singh & Moslem Azamfar (2019). Industrial Artificial Intelligence, *Computers and Society*. DOI: 10.48550/arXiv.1908.02150

MIIT (15/08/2024 a). Notice of the General Office of the Ministry of Industry and Information Technology on organizing and carrying out the collection of typical application cases of artificial intelligence enabling new industrialization,**工业和信息化部办公厅关于组织开展人工智能赋能新型工业化典型应用案例征集工作的通知**). https://www.ncsti.gov.cn/kcfw/xmsb/202408/t20240815_175395.html

MIIT (31/01/2024 b). Interpretación de las Opiniones sobre la implementación del impulso al desarrollo e innovación en las industrias del futuro, **关于推动未来产业创新发展的实施意见 解读**. Disponible en https://zwgk.mct.gov.cn/zfxxgkml/zcfg/zcjd/202401/t20240131_951103.html

MIIT + 6 (31/01/2024). Opiniones sobre la implementación del impulso al desarrollo e innovación en las industrias del futuro, **关于推动未来产业创新发展的实施意见**. Disponible en https://zwgk.mct.gov.cn/zfxxgkml/kjjy/202401/t20240131_951102.html

Nolan, Beatrice (22/01/2025). What is the 'Stargate' AI project? Disponible en https://fortune.com/2025/01/22/what-is-the-stargate-ai-project-trump-ellison-son-altman/

OICE Oficina de Información del Consejo de Estado. **国务院新闻办公室网站** (03/03/2023). **工业和信息化部部长金壮龙 : 加快推进新型工业化 做强做优做大实体经济**Jin Zhuanglong, Minister of Industry and Information Technology: Accelerate the promotion of new industrialization and make the real economy stronger, better and bigger. https://www.ncsti.gov.cn/kjdt/ztbd/2023qglh/bztd/202303/t20230303_109930.html

OpenAI & SoftBank (21/01/2025). Announcing The Stargate Project. Disponible en https://openai.com/index/announcing-the-stargate-project/

Palhares, R. M., Y. Yuan and Q. Wang, (2019). Artificial Intelligence in Industrial Systems. *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9636-9640, DOI:10.1109/TIE.2019.2916709.

Peres, Ricardo Silva, Xiaodong Jia, Jay Lee, Keyi Sun, Armando Walter Colombo & José Barata. (2020). Industrial Artificial Intelligence in Industry 4.0—Systematic Review, Challenges and Outlook. *IEEE Access*, 8, 220121–220139. DOI: 10.1109/ACCESS.2020.3042874

Pilkington, Philip (15/04/2024). Reindustrialization in the Age of Fragmentation. Disponible en https://hiia.hu/en/reindustrialization-in-the-age-of-fragmentation/

Precognize (2022) Industrial IA. https://precog.co/glossary/industrial-ai/

Qinghai Daily (22/11/2024). Chinas artificial intelligence and manufacturing industry are deeply integrated, **我国人工智能与制造业深度融合**. https://finance.sina.com.cn/jjxw/2024-11-22/doc-incwwmew4200591.shtml

Siemens (2024). Artificial Intelligence Glossary. https://www.siemens.com/global/en/company/stories/artificial-intelligence.html

Xinhua News Agency **新华社** (20/05/2024). The National Standing Committee deliberated and approved the Action Plan for Digital Transformation of Manufacturing Industry, **国常会审议通过制造业数字化转型行动方案**. Disponible en http://www.echinagov.com/news/356599.htm

CAPÍTULO CINCO

INTELIGENCIA ARTIFICIAL. EL SURGIMIENTO DE UNA NUEVA ERA EN MATERIA DE REGULACIÓN

EL DESPEGUE DE LA INTELIGENCIA ARTIFICIAL EN EL CASO BRASILEÑO. AVANCES Y RETOS REGULATORIOS POR CUMPLIR

SUMARIO

I. Introducción **II.** Dimensiones de la regulación de la inteligencia artificial: la falsa dicotomía entre la contención del daño y la aceleración del desarrollo **III**. El contexto nacional: el proyecto de ley aprobado en el Senado Federal **IV.** El contexto internacional: lecciones de la Unión Europea y las Naciones Unidas **V.** Reflexiones propositivas sobre la arquitectura regulatoria de la inteligencia artificial en Brasil **VI.** Conclusión VII. Bibliografía.

> CARLOS BAIGORRI
Presidente de la Junta Directiva de la Agencia Nacional de Telecomunicaciones (ANATEL) de Brasil

I. INTRODUCCIÓN

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías más transformadoras del siglo XXI, con aplicaciones que permean casi todos los aspectos de la sociedad, desde la salud y la educación hasta las finanzas, el transporte y más. Sin embargo, su rápido desarrollo también ha traído consigo importantes desafíos éticos, sociales y regulatorios.

En Brasil, el debate sobre la regulación de la IA tomó impulso con la presentación de un proyecto de ley en el Senado Federal, que propone un enfoque sectorial de la regulación, en lugar de la creación de un único ente regulador. Al mismo tiempo, en el ámbito internacional como en la Unión Europea (UE) y la Organización de las Naciones Unidas (ONU) se han desarrollado modelos de gobernanza que pueden servir como referencias valiosas para Brasil. Este artículo tiene como objetivo proporcionar un primer análisis de estas dimensiones, tanto nacionales como internacionales, arrojando luz sobre las posibilidades de una arquitectura regulatoria para la IA en Brasil que sea efectiva y adaptable al contexto único de nuestro país.

II. DIMENSIONES DE LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL: LA FALSA DICOTOMÍA ENTRE LA CONTENCIÓN DEL DAÑO Y LA ACELERACIÓN DEL DESARROLLO

Existe un debate frecuente sobre una aparente dicotomía entre regular la IA y fomentar su desarrollo. Muchos argumentan que la regulación de la IA puede sofocar la innovación, mientras que otros argumentan que la falta de regulación puede generar riesgos significativos. Sin embargo, esta dicotomía es en gran medida falsa. La reglamentación y el incentivo al desarrollo no son mutuamente excluyentes; de hecho, pueden ser complementarios y esenciales para garantizar que la IA se desarrolle de manera responsable y sostenible.

Atraer inversiones y desarrollar soluciones empresariales y ventajas competitivas son algunos de los beneficios que pueden derivarse del desarrollo saludable de la IA desde una perspectiva económica y financiera, por ejemplo. Por otro lado, el desarrollo no regulado de la IA también puede causar graves daños. Las amenazas a la privacidad, el sesgo algorítmico, la generación de desempleo y el uso de la IA con fines turbios e incluso delictivos requieren un enfoque responsable y cuidadoso.

En este contexto, muchos gobiernos nacionales y organizaciones supranacionales han discutido la necesidad de una regulación. La Unión Europea, por ejemplo, ha propuesto un marco regulatorio para la IA, clasificando los sistemas según el riesgo que suponen y estableciendo pautas para su uso. En Estados Unidos, existe un creciente debate sobre la necesidad de leyes que garanticen la transparencia y la rendición de cuentas en el desarrollo y la aplicación de la IA. Sin embargo, en muchos países, especialmente en aquellos países en desarrollo, existe la preocupación de que una regulación excesiva pueda impedir el progreso tecnológico y económico.

La idea de que la regulación y el desarrollo son opuestos se basa en una visión simplista del proceso regulatorio, desde la cual una regulación mal hecha afectaría la actividad económica al producir cargas regulatorias costosas e innecesarias que drenarían recursos que de otro modo estarían destinados a la innovación. Esta visión no debe ser subestimada, pero si se ejecuta de manera técnica y profesional, la regulación no será un obstáculo; por el contrario, puede servir de guía para garantizar que la innovación se produzca de forma ética y sostenible, beneficiando a todos los implicados. En este sentido, la regulación de la IA puede estimular el desarrollo de tecnologías

más seguras, transparentes y justas. Por ejemplo, los estándares que requieren la explicabilidad de los algoritmos podrían conducir a la aparición de nuevas técnicas de IA interpretables, que son más fiables y útiles en aplicaciones críticas, como el diagnóstico médico o los sistemas judiciales. Además, la regulación puede ayudar a generar confianza pública en la IA, que es esencial para masificar su adopción.

Como se ha señalado anteriormente, es importante reconocer que una regulación mal concebida puede, de hecho, impedir el pleno desarrollo de la IA. Las normas demasiado rígidas o mal adaptadas a las realidades tecnológicas pueden crear barreras donde no debería haberlas y desalentar la innovación. Por lo tanto, es crucial que la regulación sea flexible, esté basada en evidencia y se desarrolle en colaboración con expertos técnicos, empresas y la sociedad civil.

Además, la regulación debe ser proporcional al riesgo. No todos los sistemas de IA plantean los mismos riesgos. La Ley de Inteligencia Artificial de la Unión Europea, que sirvió de inspiración para el Proyecto de ley (PL) 2338/23 de Brasil, que adoptó un criterio muy similar, va en esta dirección. Es decir, un algoritmo diseñado para, por ejemplo, recomendar nuevas canciones no debería regularse de la misma manera que un sistema utilizado para guiar las decisiones judiciales. Un enfoque basado en el riesgo permite que la regulación sea focalizada y eficiente, protegiendo proporcionalmente los entornos de alto riesgo y dejando clara la libertad de innovar mucho más libremente cuando la IA se dirige a dominios considerados de bajo riesgo.

El Estado desempeña un papel central en la creación de un entorno que equilibre la regulación y los incentivos para el desarrollo de la IA. Esto incluye, entre otras iniciativas, invertir directamente y fomentar la inversión indirecta en investigación y desarrollo, promover la educación y la formación de la población en IA, facilitar y fomentar el diálogo entre los diferentes actores de la cadena de innovación, incluidas las empresas, el mundo académico y la sociedad civil, para garantizar que la regulación refleje las necesidades y deseos de la sociedad.

En este contexto, los distintos actores de la sociedad civil también tienen un papel importante que desempeñar en la regulación de la IA. La participación pública en este debate puede garantizar que las normas sean justas e inclusivas y que aborden satisfactoriamente lo que la sociedad entiende que es más importante. Además, la educación y la concienciación sobre los beneficios y riesgos de la IA son esenciales para construir una sociedad informada y comprometida y crear un entorno de seguridad institucional y paz que, en última instancia, actúe como catalizador de la inversión y generador de innovación.

Por lo tanto, la falsa dicotomía entre regular e incentivar el desarrollo de la IA es un obstáculo para el progreso responsable y sostenible de esta tecnología. En lugar de ver la regulación como una barrera, deberíamos entenderla como una herramienta para proporcionar un entorno en el que florezca la innovación. Los países que puedan equilibrar la regulación y los incentivos para el desarrollo pueden posicionarse de manera proactiva en el ecosistema mundial de innovación en IA de una manera dinámica y responsable, con el objetivo de garantizar que la IA sea una fuerza social positiva que impulse el progreso y garantice la protección de los valores fundamentales de la sociedad.

III. EL CONTEXTO NACIONAL: EL PROYECTO DE LEY APROBADO EN EL SENADO FEDERAL

Como se mencionó anteriormente, en Brasil, la regulación de la IA fue debatida recientemente en el Senado Federal, desde donde se envió a la Cámara de Diputados el Proyecto de Ley N° 2338, de 2023 (PL 2338/23), [1]que propone lineamientos para el desarrollo y uso de esta tecnología. El relator del proyecto, el senador Eduardo Gomes, ya había dicho que cree que la regulación de la IA no debe centralizarse en un solo organismo regulador. En su lugar, aboga por un enfoque sectorial, en el que los diferentes sectores están regulados por agencias específicas, guiadas por una ley general que establece principios unificadores. Este enfoque refleja la complejidad de la IA, que impregna diversos sectores de la economía y la sociedad.

En este sentido, la propuesta prevé la creación del Sistema Nacional de Regulación y Gobernanza de la Inteligencia Artificial (SIA), [2]con la coordinación asignada a la Autoridad Nacional de Protección de Datos (ANPD).[3]

1 Gaceta del Senado Federal N° 212 de 2024 - Gacetas - Actividad Legislativa - Senado Federal (https://legis.senado.leg.br/diarios/ver/118877?sequencia=372)

2 Artículo 4. X – Sistema Nacional de Regulación y Gobernanza de la Inteligencia Artificial (SIA): ecosistema regulatorio coordinado por la autoridad competente cuyo propósito principal es promover y asegurar la cooperación y armonización con otras autoridades sectoriales y entes reguladores, sin vínculo de subordinación jerárquica entre ellas, y otros sistemas nacionales para la plena implementación y cumplimiento de esta Ley en todo el territorio nacional, con seguridad jurídica;

3 Artículo 45. Se autoriza al Poder Ejecutivo a establecer el Sistema Nacional de Regulación y Gobernanza de la Inteligencia Artificial (SIA).
Parágrafo 1° - La SIA está compuesta por:
I – la Autoridad Nacional de Protección de Datos (ANPD), autoridad competente que coordinará la AIE;
II – las autoridades sectoriales;

La ausencia de un regulador único no implica una falta de coordinación. El ponente propuso la creación de un organismo central encargado de establecer las directrices básicas y armonizar las reglamentaciones sectoriales. Este organismo tendría un rol estratégico, asegurando que los diferentes organismos actúen de manera coherente y en línea con los principios establecidos en la ley. Este enfoque híbrido, que combina un órgano central de coordinación con organismos sectoriales de ejecución, parece ser una solución conciliadora en el sentido de poner fin a las disparidades entre los distintos sectores de la sociedad. La Agencia Nacional de Telecomunicaciones (Anatel) aparece como una de las probables instituciones que conformarían la SIA como autoridad sectorial.

Sin embargo, la implementación de esta arquitectura regulatoria presenta desafíos significativos. Una de las principales dificultades es armonizar las normativas sectoriales. Además, la regulación debe ser lo suficientemente flexible como para seguir el ritmo acelerado de la innovación tecnológica, sin comprometer la seguridad jurídica y la protección de los derechos fundamentales, lo que requeriría una coordinación constante. La participación de diversas partes interesadas, incluidos el gobierno, el sector privado, la academia y la sociedad civil, será crucial para superar estos desafíos y construir un marco regulatorio legítimo y efectivo.

IV. EL CONTEXTO INTERNACIONAL: LECCIONES DE LA UNIÓN EUROPEA Y LAS NACIONES UNIDAS

En el ámbito internacional, la regulación de la IA ha sido objeto de un intenso debate, con diferentes países y organizaciones supranacionales adoptando diferentes enfoques.

ORGANIZACIÓN DE LAS NACIONES UNIDAS

Es relevante destacar la iniciativa de la Organización de las Naciones Unidas (ONU), que estableció un "Grupo de Trabajo Interinstitucional" sobre IA, copresidido por la Unión Internacional de Telecomunicaciones (UIT) y la UNESCO. Este grupo tiene como objetivo promover la cooperación internacional y el desarrollo de directrices globales para la gobernanza de la IA. El compromiso de la UIT y la UNESCO refleja la naturaleza multisectorial y transnacional de la IA, que requiere un enfoque integrado y colaborativo. La experiencia de la ONU puede ofrecer valiosas perspectivas para Brasil, en particular en lo que

respecta a la coordinación entre diferentes sectores y la integración con las normas internacionales.

Más recientemente, como resultado de los acontecimientos derivados de la Cumbre del Futuro y en el ámbito del Pacto Digital Mundial (GDC), se creó una nueva estructura permanente bajo la dirección del Secretario General de las Naciones Unidas, la Oficina de Tecnologías Digitales y Emergentes (ODET)[4], que nació de la antigua Oficina del Enviado de Tecnología del Secretario General de las Naciones Unidas. Esta nueva estructura tiene como objetivo dar más eficacia a las estrategias de la Secretaría General de las Naciones Unidas frente a los nuevos retos tecnológicos disruptivos, incluida la IA. Una vez más, llama la atención el protagonismo que otorga esta nueva oficina a la Infraestructura Digital Pública, destacando la interdependencia entre la conectividad y las nuevas tecnologías emergentes.

También dentro de esta estructura funcionó la Junta Asesora de Alto Nivel para la IA (HLAB), que proporcionó al Secretario General las subvenciones relacionadas con el tema en el adelanto del Pacto para el Futuro, celebrado en septiembre de 2024 en Nueva York, en la sede de las Naciones Unidas.

A diferencia de la Unión Europea, la ONU no tiene la intención de esbozar formas de actuar y estructurar marcos regulatorios en todo el mundo, pero esta experiencia proporciona una plataforma muy interesante para los debates y muestra cómo el mundo ve los problemas relacionados en el universo de las tecnologías emergentes y sus aplicaciones facilitadoras.

UNIÓN EUROPEA

La Unión Europea (UE) se ha erigido como un actor clave en este campo, proponiendo la creación de una "Oficina de IA" compuesta por representantes de los estados miembros. Esta oficina coordina las políticas de IA a nivel europeo, asegurando la armonización de las regulaciones nacionales y la adopción de estándares comunes. El enfoque de la UE se basa en principios como la transparencia, la rendición de cuentas y la protección de los derechos fundamentales, que han servido de referencia para Brasil.

Esta Ley Europea de IA entró en vigor el 1 de agosto de 2024, dando inicio a un audaz calendario de implementación. Uno de los primeros aspectos de la aplicación nacional es la designación de autoridades. En los Estados miembros se han definido tres tipos de autoridades de acuerdo con esta Ley de IA. Aquí, una vez más, hay similitudes con lo definido en la PL 2338/23, pero también algunas diferencias significativas.

Los Estados miembros de la Unión Europea están obligados a designar o establecer tres tipos de autoridades como parte de la aplicación de la Ley de IA de la UE, preferiblemente: una Autoridad de Vigilancia del Mercado; a Autoridad Notificante y Autoridades Públicas Nacionales.

La "autoridad de vigilancia del mercado" se encargará de llevar a cabo actividades y tomar medidas sobre la vigilancia del mercado y la conformidad de los productos. Por lo tanto, esta autoridad se basa en el concepto preexistente y bien establecido de autoridades de vigilancia del mercado dentro del Derecho de la UE y será responsable de garantizar que solo los productos que cumplan con el Derecho de la UE estén disponibles en el mercado de la UE.

A su vez, la "autoridad notificante" actuará estableciendo y llevando a cabo el procedimiento para la evaluación, designación y notificación de los organismos de evaluación de la conformidad y supervisándolos. Los "organismos de evaluación de la conformidad" son organismos que llevan a cabo actividades de evaluación de la conformidad por parte de terceros, incluidos ensayos, certificaciones e inspecciones.

La autoridad notificante y la autoridad de vigilancia del mercado se denominan colectivamente «autoridades nacionales competentes». Deben operar de manera independiente, imparcial y sin sesgos, contar con los recursos técnicos, financieros y humanos adecuados, así como con la infraestructura necesaria para realizar eficazmente sus tareas designadas en la Ley de IA. Es responsabilidad de la Comisión Europea facilitar el intercambio de experiencias entre las autoridades nacionales competentes de los Estados miembros.

Por último, la "Autoridad Pública Nacional", corresponde a los Estados miembros identificar a las autoridades públicas nacionales que supervisan el cumplimiento de las obligaciones en materia de derechos fundamentales en los Estados miembros en relación con los sistemas de IA de alto riesgo. Dichas autoridades deben estar facultadas para solicitar y acceder a cualquier documentación creada o mantenida en virtud de la Ley de AI, cuando dicha documentación sea necesaria para llevar a cabo eficazmente su mandato dentro de los límites de su jurisdicción.

La Ley de IA otorga a los Estados miembros una amplia discrecionalidad con respecto a la estructura y el diseño de estos tres tipos de autoridades. En consecuencia, los Estados miembros han propuesto o designado autoridades que adoptan diversas formas. Por ejemplo, España ha establecido una Agencia Española de Supervisión de la Inteligencia Artificial (AESIA) que actúa como autoridad de vigilancia del mercado único dependiente del Departamento de Transformación Digital de España. Por el contrario, Finlandia propuso un

modelo descentralizado, nombrando a 10 autoridades de vigilancia del mercado existentes, entre ellas la Autoridad de Energía, la Agencia de Transportes y Comunicaciones y la Agencia de Medicamentos, entre otras.

Al momento de redactar este informe, aún existen varios vacíos en cuanto a las atribuciones de responsabilidad por parte de los Estados miembros, ya que se trata de una decisión que incluye una difícil negociación política al interior de cada uno de los países. Sin embargo, según una encuesta realizada por la propia Unión Europea, es posible identificar algunas decisiones ya tomadas.

Los Estados miembros deben establecer o designar autoridades competentes antes del 2 de agosto de 2025, es decir, en el momento de redactar este informe, esta obligación aún no había expirado, por lo que el panorama actual indica que ocho Estados miembros tienen propuestas legislativas pendientes o han designado una autoridad competente ("claridad parcial"), mientras que 18 Estados miembros no han designado ni establecido ninguna autoridad competente.

Por lo que respecta a las autoridades que protegen los derechos fundamentales (autoridades públicas nacionales), los Estados miembros deben publicar una lista de dichas autoridades a más tardar el 2 de noviembre de 2024. Sin embargo, solo 7 Estados miembros han publicado una lista de dichas autoridades en el momento de escribir este artículo, según la propia Comisión Europea.

CUADRO 1 Situación general de la designación de Autoridades Nacionales

Estado	Autoridades nacionales competentes	Autoridades de protección de los derechos fundamentales
Confuso	17	20
Parcialmente despejado	9	–
Claro	1	7

La única información clara sobre las Autoridades Nacionales Competentes es la de Malta, que ha determinado que su Autoridad de Innovación Digital y su Comisión de Protección de Datos actuarán conjuntamente como Autoridades Nacionales Competentes en materia de IA.

Cabe destacar, entre las señales que aún no están del todo claras, el nombramiento del organismo regulador de las comunicaciones en cuatro países: Austria, Alemania, Finlandia y Lituania. El probable nombramiento de la autoridad de ciberseguridad por parte de Italia. Y la creación de un nuevo organismo específico para la IA, probablemente en tres países: España, Polonia y Rumanía.

Como se puede observar en la tabla, siete países ya han dado a conocer la lista de autoridades que protegen los derechos fundamentales. Estas listas incluyen varias entidades, que se asemejan a las "agencias sectoriales" de la legislación propuesta por el Senado Federal brasileño. En este caso, el organismo regulador de las comunicaciones también está presente con mucha frecuencia, y cabe destacar la designación de ANACOM, en Portugal.

V. REFLEXIONES PROPOSITIVAS SOBRE LA ARQUITECTURA REGULATORIA DE LA INTELIGENCIA ARTIFICIAL EN BRASIL

A partir del análisis de los contextos nacional e internacional, este artículo busca contribuir al debate sobre una arquitectura regulatoria para la Inteligencia Artificial (IA) en Brasil, que combine la flexibilidad de las regulaciones sectoriales con la coordinación de un órgano central, tal como se propone en el PL 2338/23. Este modelo tiene como objetivo establecer lineamientos básicos y armonizar la normativa sectorial, asegurando la coherencia y eficacia del marco regulatorio. Además, este tipo de arquitectura multipartita promovería la cooperación internacional y la integración de Brasil en el escenario de gobernanza global de la IA y las tecnologías emergentes en su conjunto, condición fundamental dada la transnacionalidad del tema.

La implementación de esta arquitectura requeriría la participación de diversas partes interesadas, incluidos el gobierno, el sector privado, la academia y la sociedad civil. También sería necesario establecer mecanismos continuos de seguimiento y evaluación, que garanticen que la regulación siga siendo adaptable a los cambios tecnológicos y regulatorios. Las experiencias internacionales, en particular las de la Unión Europea (UE) y las Naciones Unidas (ONU), pueden aportar lecciones valiosas, pero es esencial que la regulación se adapte al contexto único de Brasil.

Una preocupación que destaca en la normativa europea es la exigencia de que las instituciones nacionales designadas para tratar el tema cuenten con una estructura robusta. Esta misma preocupación fue expresada en la PL 2338/23, cuando el legislador requirió al gobierno federal que provea a la Autoridad Nacional de Protección de Datos (ANPD) de los recursos necesarios -físicos, financieros y humanos- para que pueda cumplir con sus responsabilidades legales.

Esta preocupación también se observó en el momento de la decisión por parte de los distintos Estados miembros de la UE, que parecen atribuir esta responsabilidad a organismos públicos ya establecidos, con la experiencia

técnica y los medios y recursos necesarios para realizar la tarea de manera adecuada. Fue en este sentido que la Agencia Nacional de Telecomunicaciones (Anatel), dentro del legítimo debate legislativo, se puso a disposición de Brasil para que también asumiera esta responsabilidad.

Tal y como se encuentra configurada actualmente la PL 2338/23, Anatel ya comienza a prepararse para actuar dentro de sus responsabilidades y competencias legales en lo que respecta a la regulación sectorial de la Inteligencia Artificial. Sin embargo, sigue estando disponible para una eventual ampliación de su papel en el contexto nacional, si las autoridades públicas constituidas lo consideran razonable y apropiado.

En todo caso, ni la Anatel, ni la ANPD, ni ninguna otra institución estarían inmediatamente preparadas para una tarea tan compleja. Es necesario establecer y fortalecer las instancias de participación social previstas en el PL 2338/23. Una red de apoyo sectorial sólida también debe actuar constantemente. La propuesta europea, en este sentido, ayuda a abordar algunas de estas cuestiones, pero no es suficiente. Cada una de las experiencias nacionales de los Estados miembros de la UE proporcionará una orientación concreta sobre cómo hacer frente a este reto. En este contexto, destacamos la experiencia de Portugal, dado que este país tiene similitudes culturales muy cercanas con Brasil. Además, es notable la baja tasa de nombramiento de las oficinas de protección de datos personales de los países miembros como nuevos funcionarios de IA, a pesar de que la Oficina Europea de Protección de Datos es miembro observador del Consejo de IA de la UE.

Finalmente, considerando las grandes inversiones necesarias para hacer viable la ANPD como el cuerpo central del Sistema de Inteligencia Artificial (SIA) en Brasil, también valdría la pena considerar la experiencia española, que apostó por la creación de un nuevo organismo específico para la IA, un camino que parece estar siendo adoptado, al menos, también por Polonia y Rumania.

VI. CONCLUSIÓN

La regulación de la Inteligencia Artificial es un desafío complejo y multifacético que requiere un enfoque multisectorial y transnacional. Esto incluye tanto la dimensión técnica como la humana, ya que se trata de una cuestión de relevancia social que implica la protección de los derechos fundamentales, fundamentos de la vida en sociedad.

En Brasil, la oportunidad que representa la PL 2338/23 permite definir una arquitectura regulatoria que combine la flexibilidad de las regulaciones secto-

riales con la coordinación de un organismo central. Las experiencias de la Unión Europea y de las Naciones Unidas ofrecen valiosos ejemplos. El cuidado, el sentido de la responsabilidad y la urgencia con la que se debate el tema en el contexto internacional refuerzan la pertinencia del tema también a nivel nacional.

Es crucial que Brasil actúe de manera efectiva para estar en consonancia, incluso con el tiempo, con lo que están haciendo otros países. La regulación dentro de nuestras fronteras debe adaptarse al contexto brasileño, a nuestras leyes y a nuestros valores, sin dejar de lado la necesaria inclusión de Brasil como actor destacado en el escenario global de la Inteligencia Artificial.

Con la participación de diversas partes interesadas, autoridades públicas e iniciativa privada, y la adopción de mecanismos de gobernanza responsable e incentivos efectivos, Brasil puede alcanzar la posición que le corresponde tanto en el ámbito de las buenas prácticas regulatorias como en el escenario tecnológico internacional en IA. De esta manera, el país contribuirá al desarrollo sostenible e inclusivo, impulsado por una tecnología con potencial disruptivo y posiblemente revolucionario: la Inteligencia Artificial.

VII. BIBLIOGRAFÍA

Oficina de Tecnologías Digitales y Regulatorias (2025). Disponible en: https://www.un.org/digital-emerging-technologies/content/about

Senado Federal de Brasil. (2023). *Proyecto de Ley de Inteligencia Artificial*. Disponible en: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233.

Unión Europea. (2024). *Ley de Inteligencia Artificial*. Disponible en: https://artificialintelligenceact.eu/national-implementation-plans/.

Unión Internacional de Telecomunicaciones (UIT). (2023). *Cumbre Mundial de IA para el Bien*. Disponible en: https://aiforgood.itu.int/.

THE EUROPEAN AI ACT – PROMISES, PERILS & TRADE-OFFS IN A COMPLEX REGULATORY ECOSYSTEM –

SUMMARY

Introduction **I.** Key features of the AIA **I.I** The AIA's Nature, Timeline, and Status **I.II** The AIA's Risk-based approach **I.III** Relevant actors **I.IV** The AIA's Twin aims and the need for Trade-offs **I.V** Support for innovation – The need to protect startups and SMEs **II.** Analysis/discussion **II.I** Potential for global impact **II.II** Potential to respond to future technological advances **II.III** Impact on the competitiveness of Europe **II.IV** Lessons for other countries **III.** Conclusions Bibliography

> TIMO MINSSEN
LL.D., LL.Lic., LL.M.., MICL & Dipl. Jur., Professor of Law & Director of the Centre for Advanced Studies in Bioscience Innovation Law (CeBIL) at the University of Copenhagen, Timo.Minssen@jur.ku.dk

> HANNAH LOUISE SMITH
BA, M.St, BCL, DPhil, postdoctoral researcher at CeBIL, hsm@jur.ku.dk

> SEBASTIAN PORSDAM MANN
MA (Cantab.), PhD (Cantab.), DPhil (pending, Oxon.), GDL (BPP University), Assistant Professor at CeBIL, Sebastian.porsdam.mann@jur.ku.dk

INTRODUCTION

The EU's AI Act[1] (AIA) is a groundbreaking piece of legislation that represents the world's first comprehensive legal framework for artificial intelligence (AI). Initially proposed by the European Commission in April 2021, it took years of negotiation with the European Council before a final agreement was reached with the AIA entering into force on August 1, 2024. Nevertheless, many of its provisions are not yet applicable and obligations will continue to be introduced until 2026, when it will become fully applicable, with a few exceptions lasting until 2030. For example, it is only since February 2025 that the prohibitions on AI systems that pose unacceptable risks have been applicable, yet the impact of this will be muted until August 2025, when provisions relating to governance, enforcement and penalties will also become applicable.[2]

This chapter will highlight some of the AIA's features, including its objectives and provisions, as well as how it fits within a broader regulatory landscape. This provides the basis for a discussion of the promises, perils and trade-offs that the AIA is facing within a complex regulatory ecosystem and in an increasingly fierce international competition for global dominance.

1 European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (AIA)*, Official Journal of the European Union, L 311/1 (July 12, 2024).

2 European Commission. *Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AIA)*. C(2025) 884 final (February 4, 2025), Brussels. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

I. KEY FEATURES OF THE AIA

The AIA marks a seminal development in the legal oversight of AI technologies, having received approval on March 13, 2024. Although referred to as an "Act," it functions as an EU Regulation (Regulation EU 2024/1689), i.e., the strongest form of harmonizing legislation at the EU's disposal as it applies directly in all Member States without the need for further national implementation legislation. As such, once it enters into force, it will automatically become part of the legal framework of all 27 EU Member States. This places it on par with other significant regulations, such as the Digital Services Act[3], the Digital Markets Act[4], the Medical Device Regulation (MDR)[5], the Clinical Trials Regulation (CTR)[6], pending liability legislation[7], and the General Data Protection Regulation (GDPR).[8] Together, these interrelated legal instruments are collectively shaping the European landscape for the regulation of AI and digital technologies.

I.1 THE AIA's NATURE, TIMELINE, AND STATUS

Despite the three years it took from the European Commission's initial proposal in 2021 to its agreement among EU Ministers in February 2024, the EU still achieved its goal of being the first region in the world to implement comprehensive

3 European Parliament and Council. *Regulation (EU) 2022/2065 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act)* (Text with EEA relevance), PE/30/2022/REV/1, OJ L 277, (October 27, 2022):1–102.

4 European Parliament and Council. *Regulation (EU) 2022/1925 on Contestable and Fair Markets in the Digital Sector and Amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)* (Text with EEA relevance), PE/17/2022/REV/1, OJ L 265 (October 12, 2022):1–66.

5 European Parliament and Council. *Regulation (EU) 2017/745 on Medical Devices, Amending Directive 2001/83/EC, Regulation (EC) No 178/2002, and Regulation (EC) No 1223/2009, and Repealing Council Directives 90/385/EEC and 93/42/EEC* (Text with EEA relevance), OJ L 117 (May 5, 2017):1–175.

6 European Parliament and Council. *Regulation (EU) No 536/2014 on Clinical Trials on Medicinal Products for Human Use and Repealing Directive 2001/20/EC* (Text with EEA relevance), OJ L 158, May 27, 2014, 1–76.

7 European Commission, *Commission Work Programme 2025,* last modified February 12, 2025, https://commission.europa.eu/strategy-and-policy/strategy-documents/commission-work-programme/commission-work-programme-2025_en.

8 See also: Mateo Aboy, Timo Minssen and Effy Vayena, *Navigating the EU AIA: implications for regulated digital medical products.* Nature Digital Medicine 7, 237 (2024). https://doi.org/10.1038/s41746-024-01232-3

AI legislation.[9] As an EU legislative instrument, the AIA comprises of both authoritative, binding articles and recitals that provide non-binding, albeit important guidance as to the AIA's core principles and aims. These encompass: (1) ensuring the safety of AI systems and safeguarding fundamental rights; (2) fostering innovation and the adoption of AI within the EU; (3) establishing legal clarity to promote investment and advancement in AI; and (4) managing risks linked to specific applications of AI, especially those that pose substantial threats to fundamental rights.

A staggered implementation timeline and severe fines

The AIA entered into force on August 1, 2024, but its provisions will apply in different phases until 2030. Non-compliance with the AIA could lead to significant penalties, with fines reaching up to €35 million or 7% of global turnover, underscoring the importance of early preparation.[10]

This means that various actors must pay close attention to what is expected from them and when. For example, in accordance with Article 56 AIA, actors involved in general-purpose AI (GPAI) will have to be attentive to the evolving codes of practice, which are slated to apply by May 2025, and whose development is spearheaded by the AI Office with input from the AI board and a plethora of other stakeholders.[11,12] Then, in August 2025, i.e., 12 months after the AIA entered into force, various obligations for providers of GPAI models will come into effect. Member States must then also appoint competent authorities to oversee the AIA, and the Commission will conduct it first review of the list of prohibited AI practices that will then occur annually. 18 months after the AIA's entry into force, i.e., on 2 February 2026, the Commission will implement acts relating to the provisions on post-market monitoring.

9 *Ibid.*

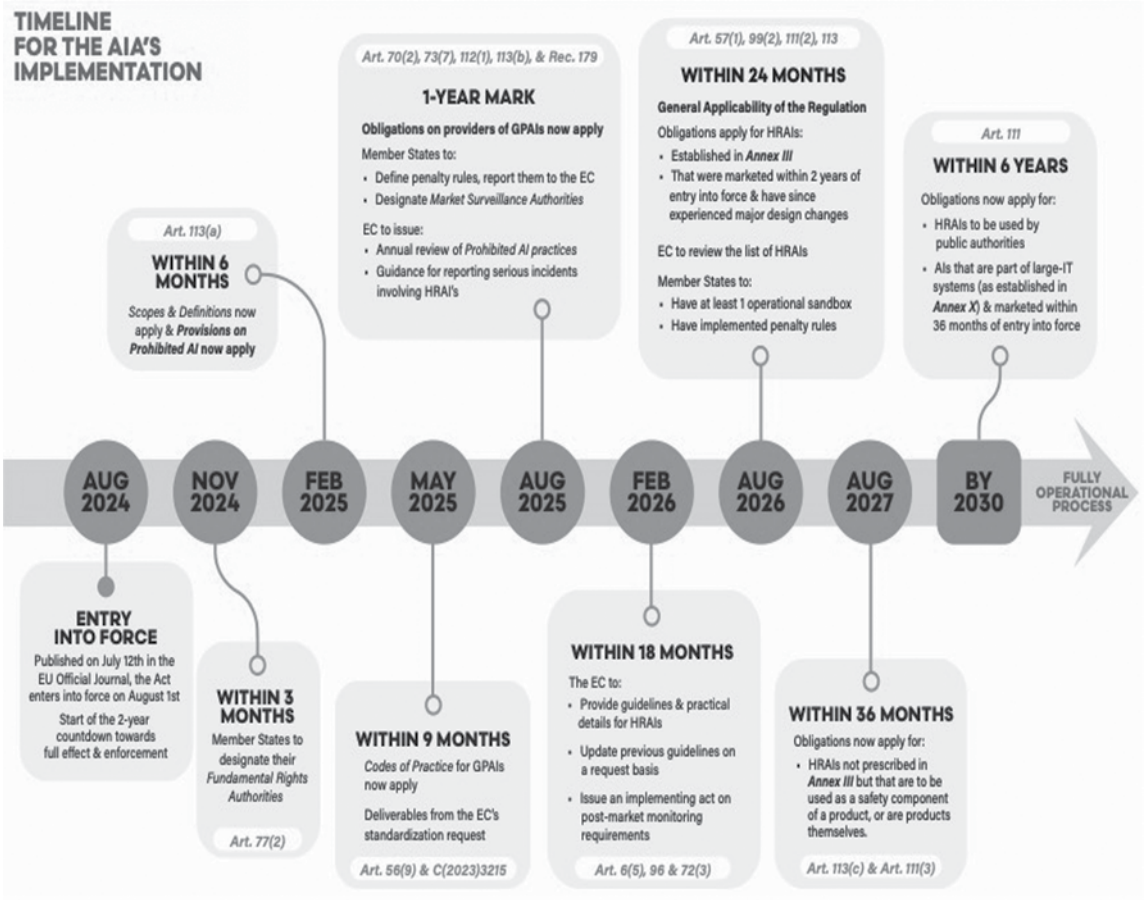
10 For an overview of the applicable penalties, see *Artificial Intelligence Act*, Art. 99.

11 For an overview on the various codes of practice under the AIA, see Jimmy Farrell and Afek Shamir, *An Introduction to Codes of Practice for the AI Act*, EU Artificial Intelligence Act, last modified October 30, 2024, https://artificialintelligenceact.eu/introduction-to-codes-of-practice/.

12 See also *Artificial Intelligence Act*, Art. 56.

Obligations for high-risk AI systems will first become applicable over the next 24 to 36 months, depending on the specific category of high-risk AI.[13] For AI systems associated with regulated digital medical products under the EU MDR/IVDR (Annex II), these obligations will start to apply 36 months after the AIA entered into force. The illustration below provides a basic overview of this implementation timeline:

FIGURE 1: Timeline for the AIA's implementation.



13 Joe Jones, *EU AIA: Next Steps for Implementation*, International Association of Privacy Professionals (January 2025), https://iapp.org/resources/article/eu-ai-act-timeline/.

Broad applicability and interactions

In contrast to sector-specific regulations that govern particular types of products, such as the EU MDR that sets requirements for the market access, safety, and efficacy of medical devices, the AIA's applicability is much broader. As AI technologies are increasingly integrated into a range of products and services that may already be regulated by sector-specific laws, one key challenge is to ensure any interactions between the AIA and other legislative instruments are harmonious. The AIA does not exist in isolation but interacts in various ways with existing legislative frameworks within the EU. Therefore, it is critical that these interactions foster an integrated regulatory environment so those developing and deploying AI understand the full range of their duties and responsibilities and individuals are aware of the protections offered to them.

Interaction with the GDPR and fundamental GDPR principles

The processing of personal data, as regulated by the GDPR, is increasingly undertaken by technologies that fall within the scope of the AIA. Therefore, the functioning and operation of these instruments cannot be evaluated in isolation. For example, both the AIA and GDPR emphasize the same principles, such as accountability and transparency. Moreover, the AIA requires that high-risk AI systems adhere to the principles of data protection as provided by the GDPR, including data minimization, purpose limitation, and storage limitation.

Nevertheless, the interplay between these instruments might not always be harmonious. For example, under Article 9 GDPR, special categories of data cannot be processed unless a specific exemption applies. In Case C-184/20[14] the EU's Grand Chamber held that if an organization could infer or deduce special categories of data, then data enabling such inferences should be treated as a special category of data. However, the AIA provides an explicit exemption from the prohibition contained in Article 9 GDPR as Article 10(5) AIA allows special categories of data to be processed where it is "strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to the high-risk AI systems" so long as appropriate safeguards are in place. Hence, providers and deployers of AI systems may have to tread very carefully and examine what may be encompassed by the term "strictly necessary" to not fall foul of the GDPR's restrictions, especially after the court's expansive interpretation of special categories of data. The European Parliamentary Research Service

14 *Case C-184/20, OT v. Vyriausioji tarnybinės etikos komisija*, ECLI:EU:C:2022:601, August 1, 2022.

(EPRS) has acknowledged the potential tension between the GDPR's restrictive approach towards processing special categories of data and the AIA's provisions on processing such data to avoid discrimination. The difficulties in reconciling the two legislative frameworks means that the EPRS is even considering whether the GDPR requires reform.[15]

Furthermore, the AIA builds on the concept of Data Protection Impact Assessments (DPIAs) that were established under Article 35 GDPR. DPIAs are required where processing is likely to result in a high risk to the rights and freedoms of individuals. However, where such an assessment is required by AI-related data processing, the European Data Protection Supervisor has expressed concerns that this interplay may not appropriately protect individuals. Organizations deploying high-risk AI systems are also required to conduct assessments to identify and mitigate risks related to safety and fundamental rights,[16] but it may be difficult to the point of near impossibility for a provider to assess all the ways its system may be used. Therefore, an AI provider's assessment may fail to correspond to the risks that arise at the deployment stage and leave individuals exposed to harm.[17]

The AIAs interaction with Product Liability and Safety Legislation

The AIA also interacts with existing (and pending) product liability[18,19] and safety legislation, including the General Product Safety Directive[20] and sector-specific regulations, such as the MDR.[21] The AIA does not establish a new liability regime

15 European Parliamentary Research Service, *Algorithmic discrimination under the AI Act and the GDPR* (February, 2025). https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2025)769509

16 *Supra* note 8.

17 European Data Protection Supervisor. *Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the Light of Legislative Developments* (October 23, 2023):11. https://www.edps.europa.eu/system/files/2023-10/2023-0137_d3269_opinion_en.pdf

18 European Commission. *Liability Rules for Artificial Intelligence.* Last modified February 12, 2025. https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en.

19 European Parliament and Council. *Directive (EU) 2024/2853 on Liability for Defective Products and Repealing Council Directive 85/374/EEC* (Text with EEA relevance), PE/7/2024/REV/1, OJ L 2024/2853, (November 18, 2024).

20 European Parliament and Council. *Regulation (EU) 2023/988 on General Product Safety, Amending Regulation (EU) No 1025/2012 and Directive (EU) 2020/1828, and Repealing Directive 2001/95/EC and Council Directive 87/357/EEC* (Text with EEA relevance), PE/79/2022/REV/1, OJ L 135 (May 23, 2023):1–51.

21 Supra note 8.

and the recent withdrawal of a legislative draft on AI liability due to a lack of consensus[22] demonstrates the seemingly intractable tension between those favoring a precautionary approach and those favoring a permissive approach. The current geopolitical climate may support an approach that promotes the EU as an engine of AI innovation, but that is no guarantee of any future prospects. However, the AIA emphasizes that existing product liability laws continue to apply to AI systems,[23] alongside any applicable liability regimes that exist within Member States. This means that manufacturers and developers of AI technologies must ensure that their products comply with safety standards and can still be held liable for certain damages caused by malfunctioning AI systems.[24] High-risk AI systems must also meet essential safety requirements, which align with existing product safety directives. This includes conducting risk assessments, ensuring human oversight, and providing necessary documentation for compliance.[25]

The AIA's interaction with the Digital Services Act and Digital Markets Act

The AIA is also part of the broader legislative landscape that includes the Digital Services Act (DSA) and the Digital Markets Act (DMA), which regulate digital services and platforms within the EU.[26]

The DSA addresses issues related to content moderation and the accountability of online platforms, which may intersect with AI technologies, and in particular GPAI, that are used for content curation and moderation.[27] For example, Article 13 AIA (Transparency and Provision of Information to Deployers), Article 27 AIA (Fundamental Rights Impact Assessment for High-Risk AI Systems), Article 50 AIA (Transparency Obligations for Providers and Deployers of Certain AI Systems,) and Article 51 AIA (Classification of General-Purpose AI Models as General-Purpose AI Models with Systemic Risk) may impose additional requirements on AI systems employed in these contexts to

22 PYMNTS, European Commission Withdraws Draft Rules on Technology Patents, AI Liability and Consumer Privacy (February 12, 2025). https://www.pymnts.com/news/regulation/2025/european-commission-withdraws-draft-rules-on-technology-patents-ai-liability-and-consumer-privacy/.

23 See e.g. the *Artificial Intelligence Act*, Recitals 9, 11, Art. 2(5), Art. 57(12), Art. 60(9).

24 *Ibid*. and *supra* note 3.

25 *Ibid.*

26 *Supra* notes 3 and 4.

27 Jordi Calvet-Bademunt and Joan Barata, *The Digital Services Act Meets the AI Act: Bridging Platform and AI Governance*, Tech Policy Press (May 29, 2024). https://www.techpolicy.press/the-digital-services-act-meets-the-ai-act-bridging-platform-and-ai-governance/.

ensure compliance with fundamental rights, ethical standards and transparency obligations.[28]

The DMA focuses on ensuring fair competition in digital markets. While the AIA primarily targets the ethical and safety aspects of AI technologies, its interaction with the DMA is spurring discussions around how uses of AI could further distort competition, foster monopolistic practices, and conflict with the EU competition rules.[29,30,31]

The AIA as a centerpiece of an emerging regulatory ecosystem

The AIA is thus the centerpiece of the EU's emerging regulatory ecosystem, which seeks to address the multifaceted challenges posed by AI. As the landscape evolves, several key themes and considerations are emerging.

As indicated above, the AIA emphasizes that AI developments must align with the EU's broader objectives of protecting fundamental rights and fostering trust in technology. This is supported by a focus on ethics that is interpreted through the lens of EU protections for individuals' rights and freedoms. This can be found in the European Commission's guidelines on trustworthy AI, which advocate for principles such as fairness, accountability, and transparency.[32] In the same vein, several recitals and articles in the AIA call for the establishment of guidelines and best practices for AI classification and development, [33,34] and encourage organizations to adopt ethical frameworks that align with EU values. Stakeholder engagement is crucial in that regard, with input from civil society,

28 *Ibid.*

29 European Union. *Consolidated Version of the Treaty on the Functioning of the European Union (TFEU)*, Articles 101 and 102. Official Journal of the European Union C 115 (May 9, 2008):94–95.

30 For further details, see, e.g., European Commission, *High-Level Group on the Digital Markets Act: Public Statement on Artificial Intelligence*, Digital Markets Act (May 22, 2024). https://digital-markets-act.ec.europa.eu/high-level-group-digital-markets-act-public-statement-artificial-intelligence-2024-05-22_en

31 Rupprecht Podszun, *From Competition Law to Platform Regulation: Regulatory Choices for the Digital Markets Act*, Economics 17, no. 1 (2023): 20220037, https://doi.org/10.1515/econ-2022-0037.

32 European Commission, *Directorate-General for Communications Networks, Content and Technology, Ethics guidelines for trustworthy AI*, Publications Office (2019). https://data.europa.eu/doi/10.2759/346720.

33 See also Article 6 of the *Artificial Intelligence Act*, Article 6.

34 *Supra* note 2.

academia, and industry being the most effective way of ensuring that diverse perspectives are considered in the development of AI technologies.[35,36]

The Importance of International Cooperation

The global nature of AI technology necessitates international cooperation and a consensus on the best governance approaches as regulatory divergences may lead to trade barriers. The EU's AIA, as the first legislative instrument governing AI technologies, stands an excellent chance of serving as a reference point for other jurisdictions considering AI regulation. The EU's emphasis on a unified regulatory approach seeks to mitigate the risks of trade barriers and burdensome compliance requirements, promoting a level playing field for businesses that, most importantly, is most aligned to its rules and preferences. Therefore, the EU should continue to build links with international organizations (e.g., OECD, ISO) in the development of global standards for AI that promote ethical practices, ensure interoperability, and protect core values.[37]

Adaptability and Future-Proofing

The rapid pace of AI innovation necessitates a regulatory framework that can adapt to changing technologies and market dynamics. Under Article 112, the AIA incorporates mechanisms for periodic review and evaluation, ensuring that it remains relevant and effective. The AIA also includes provisions for ongoing monitoring and evaluation of AI systems, as evidenced by Article 89, enabling regulators to respond to emerging risks and challenges. The tiered risk-based approach allows for flexibility in regulatory requirements, enabling innovation while maintaining safety and ethical standards. But there is much debate on whether the regulatory flexibility embedded in the AIA is enough, with some already finding its requirements to be outdated.[38]

---

35 Gail E. Crawford *et al.*, *EU AI Office Seeks Stakeholder Engagement in Drafting AI Act Implementing Rules*, Latham & Watkins LLP (July 31, 2024). https://www.lw.com/en/insights/eu-ai-office-seeks-stakeholder-engagement-in-drafting-ai-act-implementing-rules.

36 International Association of Privacy Professionals, *EU AI Act Stakeholder Map.* IAPP (May 2024). https://iapp.org/resources/article/eu-ai-act-stakeholder-map/.

37 Overseas Development Institute. *International AI Governance Must Be Truly Global*, ODI (October 25, 2023). https://odi.org/en/insights/international-ai-governance-must-be-truly-global/.

38 Sandra Wachter, *Limitations and Loopholes in the EU AIA and AI Liability Directives: What This Means for the European Union, the United States, and Beyond*, Yale Journal of Law & Technology 26, no. 3 (2024):685.

I.II THE AIA's RISK-BASED APPROACH

The AIA uses a risk-based approach and classifies AI systems into five tiers according to whether they pose unacceptable, high, systemic, limited, or minimal risks. The legal obligations differ according to the categorization to provide "proportionate and effective" rules that are tailored to the "intensity and scope" of any potential risk. Article 3(2) defines the concept of risk used in the AIA as "the combination of the probability of an occurrence of harm and the severity of that harm."

Article 5 AIA provides a list of AI systems that pose unacceptable risks and are prohibited as a result. These include AI systems that manipulate individuals to circumvent their free will (Article 5(1)(a)), make criminal risk assessments and individual predictions with no human involvement (Article 5(1)(d)), and the use of emotion recognition in the workplace or educational settings except if used for medical or safety reasons (Article 5(1)(f)), among others.

FIGURE 2: The AIA's Risk-Based Approach

Article 6 distinguishes between two types of high-risk AI systems. One type, listed in Annex I, concerns products that are already subject to ex ante conformity assessments due to existing EU harmonization legislation on health and safety. These products are AI systems that are either intended to be used as the safety component in a product or the AI system itself is the product (Article 6(1)(a)). The other type is a set of applications listed in Annex III, which provides eight

domains that cover biometrics, critical infrastructure, education and vocational training, employment, worker management, access to and enjoyment of essential private services and essential public services and benefits, law enforcement, migration, asylum, and border control management, and the administration of justice and democratic processes. These applications are lawful where certain provisions within the AIA are followed. If an AI system falls within Annex III, it is for the AI provider to assess whether the AI system does pose a significant risk to health, safety, and fundamental rights and only a positive assessment will trigger the obligations associated with high-risk systems. Article 6(3) sets out criteria that exempt AI systems from the obligations imposed upon high-risk systems. Chapter III sets out a substantial number of obligations that are imposed upon high-risk AI systems, such as those relating to risk management systems (Article 9), data governance measures (Article 10), technical documentation (Article 11), transparency provisions (Article 13), human oversight requirements (Article 14), and an appropriate level of accuracy, robustness, and cybersecurity (Article 15). Providers of high-risk AI systems must also register these systems in a public database (unless it is used for law enforcement or migration purposes), are subject to post-market monitoring by providers (Article 72) and market surveillance authorities (Article 74), and may require a fundamental rights impact assessment (Article 27). Furthermore, providers of high-risk AI systems must establish a quality management system that includes a strategy for compliance, including the need to respect fundamental rights (Article 17).

Article 50 imposes certain transparency obligations on certain AI systems. This encompasses AI systems that directly interact with individuals, produce synthetic content, or entail emotional recognition systems that are not captured by Article 5 and thus completely prohibited. These AI systems must ensure any user is informed they are interacting with an AI system, unless it is obvious to a reasonable person or where they are used to detect, prevent, investigate, or prosecute criminal offences.

A late addition to the AIA's risk classification was Article 51, which outlines the criteria for defining AI systems as general-purpose AI systems that may pose systemic risks. Recital 110 gives examples of such risks and includes "any actual or reasonably foreseeable negative effects in relation to major accidents, disruptions of critical sectors and serious consequences to public health and safety; any actual or reasonably foreseeable negative effects on democratic processes, public and economic security; the dissemination of illegal, false, or discriminatory content." Such AI systems are incongruent with a static risk classification approach, as their risk varies enormously depending upon the

prompts given by their human users. Article 53 outlines the obligations for providers of general-purpose AI models that do not pose systemic risks. The obligations include the need for up-to-date technical documentation that includes data on the model's training and testing process and results of its evaluation (Article 53(1)). Providers must also make available certain information and documentation to providers who intend to integrate the general-purpose AI model into their own AI systems (Article 53(1)(b)) unless the model is released under a free and open-source license (Article 53(2)). The public release of large language models with the ability to generate novel content, such as ChatGPT, triggered this inclusion.

The AIA implicitly classifies all other AI systems to be of minimal risk. Therefore, it imposes no obligations upon such systems but does encourage a range of stakeholders, including the European Commission, Member States, and the AI Office as the supervisory body created by the AIA, to facilitate the creation of codes of conduct (Article 95). The purpose of these instruments is to encourage voluntary compliance with the high-risk requirements, even where an AI system does not fall within the scope of Annex I or Annex III of the AIA and is thus not technically a high-risk AI system.

A regulatory approach premised upon risk is a potentially advantageous one. For example, calibrating the level of obligations based on the risk posed by AI systems helps to reduce the risk of over- and under- regulating these technologies. This provides clarity to AI providers who can accordingly tailor their research and development to align with their capacity for compliance. This will help to reduce barriers to innovation, especially for AI systems judged to not pose an unacceptable or high risk. Even for high-risk systems, prior knowledge of the obligations will allow AI producers to effectively allocate resources between developing AI systems and compliance requirements. However, the classification system does not provide absolute clarity. For example, high-risk AI systems may nevertheless be exempt from the requirements relating to such systems where they fulfil the criteria in Article 6(3). Nevertheless, these criteria have been criticized as "very vague and far-reaching" [39] that may impact negatively upon the certainty that the AIA's risk-based approach was supposed to ensure.

A risk-based regulation also aligns with other EU instruments that adopt a risk-based approach and may apply to AI systems, such as the GDPR, DSA, DMA, and the MDR. A risk-based regulatory approach to AI has also been adopted by international instruments, including the G7's International

39 *Ibid.*

Guiding Principles on AI[40] and the OECD's Framework for the Classification of AI Systems.[41] This move towards international harmonization facilitates AI's advancement by promoting a more streamlined and coherent regulatory environment. This allows AI providers to dedicate a greater proportion of their resources to research and development, rather than navigating multiple, potentially conflicting, regulatory sources.

The risk-based approach within the AIA, however, comes with various challenges. *First,* it assumes that, prior to deployment, it is possible to determine which category an AI system would fall under. The feasibility of this, given the "black box" nature of many AI systems, is questionable. Ongoing developments in the technologies and uses of AI systems, as demonstrated by the rapid rise of general-purpose AI models, may require ongoing monitoring and a re-evaluation of its classification. This promotes the protection of individuals but also undermines the clarity and certainty that a risk-based approach should foster, which may impede innovation and decrease trust in AI systems and run counter to the AIA's stated aims.

Second, there has also been significant critique of the AIA's concept of risk. As the AIA is part of the New Legislative Framework that seeks to harmonize product safety requirements, it is unsurprising that the definition resonates with a product safety conception of risk that can be measured quantitatively.[42] However, the AIA is not just about curtailing risks to individuals' health and safety but also risks posed to their fundamental rights, which are less apt to be quantified in this way. Underpinning many fundamental rights frameworks is the right to human dignity, including the EU's own EU Charter of Fundamental Rights that states in Article 1 that "human dignity is inviolable. It must be respected and protected".[43] Questions have been raised as to how far dignity can be ensured where decisions are made based on quantitative data that may not consider an individual but rather derive insights from groups with characteristics that

40 Ministry of Foreign Affairs of Japan, *Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems* (October 30, 2023).

41 Organisation for Economic Co-operation and Development, Organisation for Economic Co-operation and Development *Framework for the Classification of AI Systems*, Paris: OECD Publishing (2022).

42 Marco Almada and Nicolas Petit, *The EU AIA: A Medley of Product Safety and Fundamental Rights?*, Robert Schuman Centre for Advanced Studies Research Paper No. 2023/59 (October 18, 2023).

43 European Union, *Charter of Fundamental Rights of the European Union*, Official Journal of the European Communities, C 364/1 (December 18, 2000)

may resemble that individual.[44] Does a focus on the individual's fundamental rights neglect the risks posed to collective interests that fall outside the scope of fundamental rights yet are nevertheless vulnerable to the operation of AI systems and can entail equally deleterious consequences for individuals?[45] How can one operationalize human dignity within risk management systems and technical specifications? AI systems do not operate in a vacuum but have further ramifications and consequences, such as potentially altering our understanding of fundamental societal relationships. How can one calculate the risk posed to a general sense of social trust that underpins many social interactions considering the increased use of deepfakes that leads many to question if AI generates everything?[46]

Others have criticized the AIA's concept of risk for different reasons. Whilst the European Commission can revisit the list of high-risk systems in Annex III, the criteria listed in Article 7(2) for allowing these changes were not justified by reference to statistical data on the potential for other AI systems to cause harms captured by the AIA.[47] Furthermore, the AIA's approach to risk has fostered a very one-sided understanding of AI systems as there is an exclusive focus on the negatives associated with AI systems without a corresponding framework to articulate the benefits.[48]

As the first legislative instrument governing AI, some have lamented the AIA as a missed opportunity, particularly given the possibility of it being used as a model in other jurisdictions or to at least set the tone for future global discussions on the subject. The AIA does not necessary preclude all instances of AI systems used for predictive policing where there is human input, real-time and ex-post remote biometric identification in public spaces, and biometric

44 Carsten Orwat *et al.*, *Normative Challenges of Risk Regulation of Artificial Intelligence*, NanoEthics 18, no. 2 (2024): 11, https://doi.org/10.1007/s11569-024-00454-9.

45 Sue Anne Teo, *How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework*, Nordic Journal of Human Rights 40, no. 1 (2022): 216–34, https://doi.org/10.1080/18918131.2022.2073078.

46 Atoosa Kasirzadeh, *Two Types of AI Existential Risk: Decisive and Accumulative*. arXiv preprint arXiv:2401.07836 (2024).

47 Ljupcho Grozdanovski and Jérôme De Cooman, *On the Obsolescence of Empirical Knowledge in the Age of AI*, Rutgers Computer & Technology Law Journal 49, no. 2 (2023): 123–145

48 Martin Ebers, *Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AIA*, European Union Law Working Papers, (October 21, 2024), Stanford-Vienna Transatlantic Technology Law Forum.

categorization systems.[49] The concern is that such systems have such poor accuracy rates, known issues with generating biased outputs, and obvious implications for several fundamental rights that even the obligations placed on high-risk systems are insufficient.

I.III RELEVANT ACTORS

The AIA establishes comprehensive obligations for multiple entities across the AI value chain, collectively termed 'operators' and defined in Article 3(8) as "a provider, product manufacturer, deployer, authorized representative, importer or distributor."

The Act imposes its most stringent requirements on 'providers'. According to Article 3(3), a provider is "a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge."

Deployers, defined in Article 3(4) as "a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity," also have obligations that vary by system type. Under Article 26(1), deployers must take "appropriate technical and organizational measures" to ensure they use systems in accordance with provided instructions. For emotion recognition or biometric categorization systems, Article 50(3) requires deployers to inform exposed individuals of the system's operation.

Importers, defined in Article 3(6) as "a natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established in a third country," must verify fundamental compliance elements under Article 23(1). This includes confirming appropriate conformity procedures, technical documentation, CE marking, and authorized representative appointment. For high-risk systems, Article 23 prohibits market placement of non-compliant systems, requires proper storage conditions, mandates ten-year documentation retention, and establishes active cooperation with authorities on compliance measures.

Distributors, defined in Article 3(7) as "a natural or legal person in the supply chain, other than the provider or the importer, which makes an AI system

49 *Supra* note 38.

available on the Union market," have specific obligations for high-risk AI systems under Article 24. Before making such systems available, they must verify CE marking, EU declaration of conformity, instructions for use, and compliance of providers and importers with their labeling obligations. Article 24 further requires distributors to ensure proper storage, take corrective actions for non-compliant systems, and immediately notify authorities of risks that fall within the scope of Article 79(1).

Article 25 establishes important cross-cutting provisions affecting all actors. Any operator making substantial modifications to, or changing a high-risk AI system's intended purpose, or putting their own name or trademark on an already marketed high-risk AI system, assumes full provider obligations for that system. Where this happens, the entity previously holding provider obligations ceases to do so, except where the previous provider has clearly specified that their AI system is not to be changed into a high-risk AI system.

I.IV THE AIA's TWIN AIMS AND THE NEED FOR TRADE-OFFS

The EU's legislative authority is constrained by certain principles enshrined in its founding treaties, including the need for a specific legal basis within these treaties. These legal bases provide an ongoing "steer" during the legislative process and, ultimately, influence the legislative instrument itself. The AIA rests on two such bases: Article 114 TFEU and Article 16 TFEU.[50]

Article 114 TFEU enables the EU to adopt harmonizing measures across its Member States to establish and maintain the functioning of the internal market.[51] Harmonization operates to prevent fragmentation across the single market as Member States develop their own, potentially diverging, laws. Here, the consistency created by harmonization also operates to ensure legal certainty for those who produce, deploy, and use AI systems across the EU. Article 16 TFEU justifies the EU's legislative competence in the areas of personal data protection and the free movement of data. This legal basis aligns with the EU Charter of Fundamental Rights that upholds both a right to privacy and a separate right to the protection of personal data and highlights its resonance with the GDPR.

50 European Commission. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.* COM(2021) 206 final (April 21, 2021).

51 See *TFEU*, Article 114(1).

These twin aims are reflected in Article 1(1) of the AIA, which defines the AIA's scope as the improvement of the functioning of the internal market whilst ensuring the protection of fundamental rights through the uptake of human-centric and trustworthy AI. A focus on the internal market explains the AIA's inapplicability to AI systems used exclusively for military, national security, or scientific research and development. (Articles 2(3) and 2(6)). The AIA harmonizes the rules relating to CE marking and compliance standards to integrate AI systems into the broader product safety legislative framework to provide legal certainty for producers and providers of AI systems and demonstrate AI systems' safety and trustworthiness to AI users.

These conditions are vital to the EU's efforts to promote innovation and growth within the internal market. In keeping with its fundamental rights underpinning, the AIA expands the concept of relevant risks within the existing product safety framework to incorporate risks into fundamental rights (Article 65(1)). This means developers of high-risk AI systems must conduct a pre-market assessment that incorporates potential impacts on fundamental rights and mitigation strategies (Article 6(2)). Furthermore, it requires deployers of high-risk AI systems to undertake fundamental rights impact assessments if they are governed by public law or are private entities providing public services or acting on behalf of public bodies (Article 27) and to consider the impact on fundamental rights within their risk management systems (Article 9) and data governance systems (Article 10).

Recognizing the role and relevance of fundamental rights to the legislation of AI systems is a positive development because it expands the understanding of relevant risks to better ensure AI systems can only be used for beneficial purposes. The weaving of the economic imperative of the single market and the ethics of AI shows how the development of legislation in this space is concerned with the integration of these aims, reflecting the EU's commitment to both these goals. This linkage is acknowledged in the wider corpus of EU materials concerning the governance of AI where respect for fundamental rights is not just seen as essential in itself, but also as a means of facilitating public acceptance and the competitive advantage of EU-based companies.[52] Within the AIA itself, the integration of the two aims is demonstrated in the drive for standards, a key technical means for demonstrating compliance with the AIA,

52 European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Building Trust in Human-Centric Artificial Intelligence*, COM(2019) 168 final (April 8, 2019).

to reflect fundamental rights considerations, and the introduction of a specific fundamental rights impact assessment.

In their work on the EU's Digital Single Market, Neergaard and De Vries argue that laws pertaining to the single market could act as a "normative corridor" to accommodate other objectives, including fundamental rights protections.[53]This metaphor implies a shared collective understanding of the internal market with common norms and standards that also shape the approach to fundamental rights, whilst still leaving some room for differences.[54] Within the AIA, the concept of human-centric and trustworthy AI may operate to provide a foundation for common norms that apply to the EU's economic imperative and fundamental rights protections. This would support a harmonious interpretation of the AIA that supports its twin aims.

However, the view that the AIA's twin aims may operate harmoniously in tandem is not universally held. First, some have questioned the appropriateness of a product safety framework for the protection of fundamental rights. Whilst the AIA expands the concept of risk to encompass risks to fundamental rights, the definition of risk within the AIA may fail to capture the types of risks AI systems pose to fundamental rights. Understandings of risk through a product safety lens may neglect sociopolitical risks and risks arising from the broader context of AI use, rather than just the AI system.[55]

Second, there are also potential gaps and grey areas where fundamental rights protections may fall away. The impact of AI has been particularly felt within the fields of medicine and healthcare.[56] Whilst the risks to patients are numerous and potentially severe,[57] in contrast to the internal market, the EU has a much more limited competence to legislate over health-related matters to the potential detriment of patients. Furthermore, while the AIA references fundamental rights, the legal basis concerns the protection of personal data. This basis is unlikely to cover all the risks posed to patients in a healthcare context, such as infringements on their rights to access healthcare, non-discrimination,

53 Ulla Neergaard and Sybe de Vries, *The Interaction between Free Movement Law and Fundamental Rights in the (Digital) Internal Market*, Renforce Working Paper 1-2023 (2023): 1–30.

54 Cathleen Kantner, *What Is a European Identity? The Emergence of a Shared Ethical Self-Understanding in the European Union*, EUI RSCAS Working Papers 2006/28 (2006).

55 *Supra* note 42.

56 *Supra* note 9.

57 W. Nicholson Price II and I. Glenn Cohen, *Locating Liability for Medical AI*, DePaul Law Review 73, no. 2 (2024): 339–368.

and freedom from inhumane treatment. This suggests the AIA may promote the wider use of medical AI within the EU that fully promotes the EU's economic aims but does not offer the same level of fundamental rights protections.[58] The requirement to conduct a fundamental rights impact assessment only applies to certain deployers that are either public bodies or private entities providing a public service or acting on behalf of a public body. This potentially leaves a range of AI systems not subject to this requirement, including those relating to employment and certain aspects of the financial sector.[59]

Finally, the incorporation of fundamental rights into a product safety framework may lead regulators to underestimate the need to develop new capabilities. Under Article 40, fundamental rights may fall within scope of the harmonized standards that are developed by European standardization organizations (ESOs). These are private, albeit non-profit, entities that now play a critical role in translating fundamental rights into technical standards. This plays a vital role in operationalizing the AIA, despite their lack of democratic legitimacy and the fact that most stakeholders within these groups are from industry.[60] This development in their role raises questions over whether they have sufficient expertise and appetite to appropriately respond to the challenges that AI may pose to fundamental rights. The role of ESOs in operationalizing fundamental rights protection under the AIA has also been criticized for lacking the transparency and openness of the EU legislative process. Despite the essential role these standards will play in ensuring AI systems do not impact fundamental rights to the detriment of individuals, there is no guarantee of free and easy access to the standards once they are agreed.[61] This is not the only source of opacity that runs counter to the effective operation of fundamental rights. Such opacity triggers uncertainty that undermines the efforts to create human-centric and trustworthy AI that promotes the economic aims of the EU and fundamental rights.

58 Hannah van Kolfschooten, *EU Regulation of Artificial Intelligence: Challenges for Patients' Rights*, Common Market Law Review 59, no. 1 (2022): 81–112.

59 *Supra* note 38.

60 Michael Veale and Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*, Computer Law Review International 22, no. 4 (2021): 97–112, https://doi.org/10.9785/cri-2021-220402.

61 Marion Ho-Dac, *Considering Fundamental Rights in the European Standardisation of Artificial Intelligence: Nonsense or Strategic Alliance?*, arXiv (January 23, 2024).

I.V SUPPORT FOR INNOVATION – THE NEED TO PROTECT STARTUPS AND SMEs

The relationship between regulation and innovation is complex.[62] While regulation can act as a powerful stimulus for innovation by creating market certainty and driving improvements in products and processes, it can simultaneously impose compliance costs and restrictions that may discourage innovative activities. The potential impacts of regulation on innovation are particularly salient for emerging technologies, like AI, as well as for small and medium-sized enterprises (SMEs), who are less able to bear the costs of regulatory compliance. Experience with previous regulatory interventions suggests that the impact of regulation on innovation is highly context-dependent and varies significantly based on factors including timing, stringency, flexibility, and certainty.[63]

Several provisions of the AIA could potentially constrain innovation through both direct and indirect mechanisms. First, the Act's stringent requirements for high-risk AI systems impose substantial financial compliance costs. The European Commission estimates compliance costs for providers at approximately €6,000–€7,000 per AI system, with deployers facing additional human oversight costs of €5,000–€8,000 per year.[64] For systems requiring third-party conformity assessment, providers face further verification costs of €3,000–€7,500, assuming they already have a Quality Management System (QMS) in place. These costs could be significantly higher – with estimates of ca. €193,000–€330,050 for larger enterprises and €144,350–€247,150 for SMEs - if a new QMS needs to be established.[65] Second, the extensive testing and validation requirements for high-risk AI systems under Articles 9 and 10 may significantly extend development cycles and increase costs. While the European Commission estimates that some 15% of AI systems would fall in this category, a survey conducted by the appliedAI initiative found that at least a third of AI startups would classify their products as high-risk AI systems.[66]

---

62 Jacques Pelkmans and Andrea Renda, *Does EU Regulation Hinder or Stimulate Innovation?*, CEPS Special Report No. 96 (November 19, 2014), Brussels: Centre for European Policy Studies.

63 *Ibid.*

64 Meeri Haataja and Joanna J. Bryson, *What Costs Should We Expect from the EU's AIA?*, SocArXiv (2021).

65 *Ibid.*

66 appliedAI Initiative, *AIA Impact Survey: Pan-European Survey of AI Startups and VCs about the Impact of the AIA*, AppliedAI.De, (December 12, 2022). https://www.appliedai.de/assets/files/AI-Act-Impact-Survey_Slides_Dec12.2022.pdf.

Beyond these direct financial impacts, the Act may constrain innovation through several indirect mechanisms. The transparency and interpretability requirements in Article 13 present technical challenges, since explainability methods have sometimes been found to be imprecise, inconsistent, and vulnerable to attacks.[67] Meeting these requirements may therefore force providers to favor simpler, more interpretable models over potentially more effective but less transparent approaches. The lack of clear definitions for key terms like 'sufficient transparency' and 'interpretability' creates legal uncertainty, and the extensive documentation requirements may divert resources away from R&D activities, especially affecting SMEs and startups with limited staff and budgets.

While there are thus substantial reasons to worry about the effect of the AIA on innovation, the Act also incorporates several measures specifically designed to support innovation. Most notably, Articles 57-63 establish a comprehensive framework centered around regulatory sandboxes and other support mechanisms. Regulatory sandboxes are controlled environments where companies can develop and test innovative AI systems under regulatory supervision.[68] According to Article 57, Member States must ensure that at least one AI regulatory sandbox is established at national level by August 2026. These sandboxes serve multiple objectives laid out in Article 57(9): improving legal certainty for AI development, supporting best practice sharing through cooperation between authorities, fostering innovation and competitiveness, facilitating market access for AI systems (particularly from SMEs and startups), and contributing to evidence-based regulatory learning. The sandbox framework includes several specific provisions to ensure accessibility and effectiveness, with Article 58(2) mandating broad and equal access through transparent selection criteria and requiring participation to be free of charge for SMEs and startups.

67 Luca Holst *et al.*, *The Impact of the EU AIA's Transparency Requirements on AI Innovation*, in Proceedings of the 19th International Conference on Wirtschaftsinformatik, Würzburg, Germany (2024).

68 Thomas Buocz, Sebastian Pfotenhauer, and Iris Eisenberger, *Regulatory Sandboxes in the AIA: Reconciling Innovation and Safety?*, Law, Innovation and Technology 15, no. 2 (2023): 357–389, https://doi.org/10.1080/17579961.2023.2245678.

While the AIA itself does not define SMEs, startups, and microenterprises, it builds upon established EU definitions, particularly Commission Recommendation 2003/361/EC.[69] The AIA's provisions for these entities thus likely follow these established thresholds, though notably the term 'startup' remains undefined in both instruments.

Article 59 lays out provisions for further processing of personal data within regulatory sandboxes for specific public interest purposes, including public health, environmental protection, and energy sustainability. These provisions contain multiple safeguards intended to ensure compliance with the GDPR, further supporting innovation by providing a clear path for complying with two distinct but relevant regulatory instruments. This is complemented by Article 60's framework for real-world testing outside sandboxes, subject to strict safeguards including market surveillance authority approval and informed consent requirements detailed in Article 61.

Article 62 mandates that Member States provide SMEs and startups with priority access to regulatory sandboxes and establish specific support measures including awareness-raising, training activities, and dedicated communication channels. The article also stipulates that the needs and interests of SMEs shall be taken into account when the fees for conformity assessment under Article 43 are set, "reducing those fees proportionately to their size, market size and other relevant indicators." Finally, Article 62 requires the AI Office to develop standardized templates and maintain a single information platform to ease compliance burdens. Article 63 further allows microenterprises to comply with quality management system requirements in a simplified manner, though without compromising on core obligations regarding high-risk AI systems.

Article 2 of the AIA includes several important exemptions related to research and development. Most notably, the Act "does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development." Additionally, any

69 European Commission, *Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises* (Text with EEA relevance) (notified under document number C(2003) 1422) OJ L 124, (20 May 2003):36-41. This Recommendation defines: -Microenterprises as having fewer than 10 employees and annual turnover/balance sheet total not exceeding €2 million, - Small enterprises as having fewer than 50 employees and annual turnover/balance sheet total not exceeding €10 million, - Medium-sized enterprises (completing the SME category) as having fewer than 250 employees and annual turnover not exceeding €50 million and/or balance sheet total not exceeding €43 million.

"research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service" is exempted. What this will mean in practice is, however, not easy to predict.[70] In particular, it is unclear what will count as development for the 'sole purpose' of research and development.

Much of contemporary research and development, in AI as well as more broadly, is a hybrid endeavor mixing private and public actors and scientific and commercial motivations.[71] Similarly, the Act's explicit exclusion of real-world testing from this exemption introduces questions related to the ability to neatly separate pure research from ongoing development efforts.[72] Finally, scholars have identified potential mechanisms for 'regulatory arbitrage' whereby AI system providers might structure their activities to fall within the research exemption while actually pursuing commercial ends.[73]

While there are thus clear risks that the AIA could constrain innovation through both direct compliance costs and indirect effects, regulation can also facilitate innovation by creating clear market signals, fostering predictable investment environments, and driving improvements in products and processes to meet new standards.[74] While compliance costs and technical constraints may create headwinds for certain types of AI development, particularly among resource-constrained SMEs, the Act's establishment of clear standards and support mechanisms could help create the predictability and trust necessary for AI innovation.

70 Sebastian Porsdam Mann, I. Glenn Cohen, and Timo Minssen, *The EU AIA: Implications for U.S. Health Care*, NEJM AI 1, no. 11 (October 2024). https://ai.nejm.org/doi/pdf/10.1056/AIp2400449.

71 Liane Colonna, *The AIA's Research Exemption: A Mechanism for Regulatory Arbitrage?* in YSEC Yearbook of Socio-Economic Constitutions 2023: Law and the Governance of Artificial Intelligence, ed. Andreas Moberg and Eduardo Gill-Pedro (Cham: Springer Nature Switzerland, 2024)

72 *Ibid.*

73 *Ibid.*

74 Shuai Shao *et al.*, *Environmental Regulation and Enterprise Innovation: A Review*, Business Strategy and the Environment 29, no. 3 (2020): 1465–1478, https://doi.org/10.1002/bse.2446.

II. DISCUSSION

Our above analysis revealed that the AIA is a very sophisticated and ambitious legislation. But what promises, perils and trade-offs is the AIA now facing within a complex regulatory ecosystem and in an increasingly fierce international competition? Will the AIA really serve as a regulatory north star both within the EU and beyond, or does its intended operation risk being undermined by implementation challenges and concerns? These are complex questions which need to be discussed from a variety of angles.

II.1 POTENTIAL FOR GLOBAL IMPACT

Like other major EU Regulations, such as the GDPR, Digital Services Act, and Digital Markets Act, the AIA forms part of the growing EU regulatory infrastructure aimed at protecting fundamental rights while promoting responsible innovation.

While the Act will therefore have a direct impact on the legislation of its Member States, its effect will also be felt beyond EU borders. This is due, in part, to its scope, as defined in Article 2, which stipulates that the Act applies to any provider placing AI systems on the EU market or whose systems' outputs are used in the EU, regardless of their location. The AIA can therefore create obligations and responsibilities for deployers, importers, and distributors of AI systems, as well as manufacturers incorporating AI into their products irrespective of their location. This extraterritorial reach means that any organization whose AI systems touch the EU market must comply with the Act's requirements. Notably, the Act applies even when only the output of an AI system is used in the Union, a particularly far-reaching provision that could capture many systems operated entirely outside EU territory.

This broad territorial scope reflects a well-documented phenomenon in EU regulation known as the "Brussels Effect", whereby EU regulations can become de facto global standards through market mechanisms.[75] This is, in part, because global companies must meet EU standards to launch their products in this substantial market and, rather than adhere to multiple compliance regimes in different territories, companies may prefer to apply the EU rules to all their operations. Furthermore, the adoption of a unified approach may be commercially attractive where this adherence to stringent regulations can signal

75 Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (New York: Oxford University Press, 2020), https://doi.org/10.1093/oso/9780190088583.001.0001.

quality and trustworthiness, and because it may be easier to design products for compliance with the strictest standards rather than maintain separate products for different markets. The GDPR illustrates this effect well - while technically only governing data protection in the EU, its requirements have shaped global privacy practices as companies found it more efficient to apply GDPR standards universally rather than maintain different standards for different markets. We may also see a *de jure* Brussels Effect where countries outside of the EU may adopt regulations very similar to the AIA, either due to commercial lobbying from businesses who would prefer a unified regulatory regime to reduce compliance costs or through the EU's own political influence.[76]

Whether the AIA will lead to a Brussels Effect is not a straightforward question. Neither is the distinct, albeit related, question of whether such an influence is desirable. On the one hand, some have pointed out that the Act's complexity – compared to the GDPR's relatively more straightforward principles and obligations – makes wholesale adoption by other jurisdictions unlikely.[77] However, others have noted that the Act seems to broadly satisfy the five criteria originally laid out by Bradford as necessary for a Brussels Effect:[78] significant market power (450 million EU consumers), regulatory capacity (as demonstrated by past digital regulations), stringent standards (particularly for high-risk and general-purpose AI), inelastic targets (companies cannot easily avoid compliance), and the regulation targeting non-divisible conduct (as creating separate AI systems for different markets is likely to be impractical or costly).

Yet even if the Act does produce a Brussels Effect, Almada and Radu warn that this could be accompanied by what they term a "Brussels Side Effect" - a diminution in fundamental rights protection globally.79 While the EU's competence to regulate fundamental rights –founded upon the Charter of Fundamental Rights and Article 16 TFEU specifically enabling data protection regulation - Almada and Radu argue that the Act's product safety framework may inadequately capture certain fundamental rights dimensions. This limitation could lead to the global adoption of the Act's mechanisms actually reducing

76 *Ibid.*

77 Christopher T. Marsden and George Christou, *Artificially Intelligent Regulation: Global Norms, International Political Economy and the Brussels Effect*, in Proceedings of the 2024 International Conference on Artificial Intelligence and Regulation (2024)

78 Charlotte Siegmann & Markus Anderljung, *The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market*, Centre for the Governance of AI (Aug. 2022). https://cdn.governance.ai/Brussels_Effect_GovAI.pdf

79 Marco Almada and Anca Radu, *The Brussels Side-Effect: How the AIA Can Reduce the Global Reach of EU Policy*, German Law Journal (2024).

overall fundamental rights protection by promoting an incomplete framework for protecting rights. Thus, on their analysis, a substantial Brussels Effect is not necessarily to be desired – due to the potential for this to undermine, rather than advance, the EU's regulatory policy goals.

The Act's global impact will also be shaped by the alternative regulatory approaches that are emerging worldwide. The United States does not, so far, have a comparatively systematic regulatory framework for AI. President Biden's Executive Order on AI[80], which was binding only on executive branch staff and agencies, emphasized safety testing for advanced AI systems while preserving innovation. However, this Order has now been revoked and replaced by Trump's executive order,[81,82] which may undermine arguments of the existence of a Brussels Effect. Trump's Executive Order emphasizes the US's desire to "act decisively to retain global leadership in artificial intelligence" and removes all references to guardrails rooted in ethical principles that resonate with the EU's governance approach. China has taken a more comprehensive approach through various regulations including the Algorithm Recommendation Management Provisions, though these reflect different priorities around social control and national security. At the international level, both the OECD AI Principles and UNESCO's Recommendation on the Ethics of AI have established influential soft law frameworks emphasizing human-centric AI development, though these do not have directly binding legal force. Most notably, the Council of Europe is developing its own AI Convention, which takes a fundamentally different approach centered on human rights, democracy, and the rule of law rather than product safety.[83]

---

80 The White House. *Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.* Statements and Releases (October 30, 2023). https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

81 Joseph R. Biden Jr., *Executive Order 14141: Advancing United States Leadership in Artificial Intelligence Infrastructure*, Federal Register 90, no. 5469 (January 17, 2025): 5469–5489.

82 Donald J. Trump, *Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence*, Federal Register 90, no. 5469 (January 23, 2025): 5469–5471. https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence.

83 Hannah van Kolfschooten and Carmel Shachar, *The Council of Europe's AI Convention (2023–2024): Promises and Pitfalls for Health Protection*, Health Policy 138 (December 2023): 104935, https://doi.org/10.1016/j.healthpol.2023.104935.

Given these dynamics, the AIA's global impact is likely to be significant but more nuanced than previous instances of the Brussels Effect. Much depends on the extent to which its technical requirements will diffuse through market mechanisms, as well as on how key provisions and exemptions are interpreted. Ultimately, the extent of the AIA's global influence will depend largely on how effectively the Act balances competing demands, such as protecting fundamental rights while fostering innovation, and providing clear rules while maintaining flexibility. Given the rapid pace of both the underlying technology as well as global regulatory responses, the ultimate impact of the AIA is not predictable at present.

II.II POTENTIAL TO RESPOND TO FUTURE TECHNOLOGICAL ADVANCES

There are several aspects that can be considered to evaluate the extent to which the AIA may be able to respond to future technological advances. The *first* is to consider the ability of the provisions to respond to novel and potentially unexpected technological developments. Will the risk-based system be able to accommodate technological advancements? Do the provisions allow the relevant actors to gather high-quality information to determine how the AIA operates in relation to technological advances? The *second* is related to the actors involved in the implementation, use, and enforcement of the provisions – how far are they able and willing to respond to future technological developments within the AIA? *Finally*, the AIA does not operate in a vacuum so the potential to respond to future technological developments will also depend on its ability to form a part of a wider coherent regulatory regime.

With regard to the *first* aspect, the provisions show some potential to adapt to future technological advances. The AIA uses a broad, internationally recognized definition of AI that should help ensure that it captures future AI approaches and systems. Furthermore, the AIA's focus is often on broad use cases and applications that further support its ability to capture future technological advances where they are deployed in areas covered by the AIA. In addition, Article 112 outlines the Commission's ability to assess the need for it to make amendments to Annex III, which provides one source of high-risk AI systems, and Article 5, which sets out prohibited practices. The Commission will undertake the assessment annually and report its findings to the European Parliament and Council. Indeed, Article 112 sets out detailed procedures for the review and evaluation of the AIA on several points relevant to the AIA's ability to respond to future technological

developments, including amendments to provisions concerning the definition of high-risk AI systems, AI systems requiring additional transparency measures, and the effectiveness of the supervision and governance system (Article 112(2)). Nevertheless, many of the mechanisms for evaluation and review listed in Article 112 are only required to take place every four years, which may lead to these evaluations lagging behind technological developments and responding too late. Moreover, some provisions have already been criticized for being out-of-date and not responding to current technological trends, let alone future ones. One of the means of defining general-purpose AI is threshold of 10^{25} FLOPS (floating point operations). The AIA assumes that AI models that use more than this amount of computation are a general-purpose AI model, unless a developer can demonstrate otherwise (Article 52(2)). The assumption that computing power is a defining feature of when a general-purpose AI system will lead to systemic risks has been criticized as "unconvincing"[84] and efforts to optimize AI models further undermine the relevance of the 10^{25} FLOPS threshold. This prompts concern that AI developers will focus on making AI models that fall below this threshold, rather than developing ones that are less risky. The recent release of models well under this threshold which nevertheless demonstrate state-of-the-art performance, such as DeepSeek-R1 and GPT-o3-mini, further challenges the relevance of this threshold as a proxy for systemic risk.

Now we turn to the *second* aspect, i.e., the ability of the actors involved in the application of the AIA to respond appropriately to technological developments relevant to the AIA's operation. These relevant actors include AI providers and deployers that implement the measures required to reduce the risk posed by AI systems, ESOs that develop appropriate standards to guide AI development in a way that fosters the competitiveness of the single market whilst respecting fundamental rights, and public institutions tasked with the monitoring of these actors and any necessary enforcement action. The more one focuses on technical standards to ensure the AIA responds to technological advances, the more regulatory choices are depoliticized and discussions as to values and norms, led by public institutions, are replaced with statistics and technical facts by private institutions. This leads to public institutions being largely in the dark about the day-to-day operations of the AIA, a state of affairs that becomes very problematic if technological advances then require them to intervene. The AIA contains provisions that enable public authorities to receive information on conformity assessments for high-risk AI systems (Article 21), on the operation

84 *Supra* note 38.

of general-purpose AI models (Articles 52-55), information on how far high-risk AI systems impact fundamental rights (Article 77), and to supervise, investigate, enforce, and monitor general-purpose AI risks (Section 5, Chapter IX). However, the ability to use that information to respond to potential technological developments is a distinct matter with many questioning the capability of public authorities involved in the monitoring and enforcement of the provisions,[85] both of which are a valuable source of information on how to respond to novel AI technologies.

The *final* aspect concerns how well the AIA will interact with the wider regulatory landscape to ensure it responds to all relevant risks without the duplication of efforts or resources across different, albeit related, regulatory schemes. The AIA forms part of the New Legislative Framework and interacts with other EU legislation that governs digital technologies including the General Data Protection Regulation, the Digital Service Act, the Digital Markets Act, the European Health Data Space, and the Medical Devices Regulation. Questions remain as to how far these instruments will co-exist on the broader regulatory landscape concerning data, technology, and data-driven technology companies. Questions as to how far a GDPR-compliant data protection impact assessment will align with an AIA-required fundamental risks impact assessment already demonstrates the scope for duplication or, worse, allowing risks to fall between the cracks of different regulations.

II.III IMPACT ON THE COMPETITIVENESS OF EUROPE

The intentions of the AIA, in seeking to minimize the risks of AI and maximize the potential benefits in a way that respects fundamental values and broader societal goals, are commendable.

However, as stressed in the so-called Draghi Report[86], overly excessive and complex regulations also pose significant implications and risks for the technological competitiveness of the EU in the rapidly evolving global AI landscape.

---

85 *Supra* note 38.

86 Mario Draghi, The Future of European Competitiveness (September 2024), available at: https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en ; see also European Commission, A Competitiveness Compass for the EU, Brussels, 29.1.2025 COM(2025) 30 final, available at: https://commission.europa.eu/document/download/10017eb1-4722-4333-add2-e0ed18105a34_en .

Recent examples, such as the call by the EU Parliament to reform the MDR[87] the recent withdrawal of forthcoming EU legislation on AI liability,[88] plans to simplify and streamline the GDPR[89] and the AIA[90], and the so-called OMNIBUS packages[91] have demonstrated that it is a very delicate task to simultaneously balance the risks of over- and under-regulation, protect citizens' rights and other EU values, and enhance the competitiveness of the EU.[92] In other words, it is not easy to find a "regulatory sweet spot" that promotes the beneficial application of AI innovation and reduces the potentially harmful risks that this technology entails.

One of the primary concerns regarding the AIA is the potential for increased regulatory burdens on businesses, particularly startups and SMEs. As previously discussed, the classification of AI systems into risk categories—ranging from minimal to unacceptable risk—imposes varying levels of compliance requirements. For instance, high-risk AI applications will face stringent obligations, including robust documentation, risk assessments, and ongoing monitoring. This could create a challenging environment for smaller entities that may lack the resources to navigate complex regulatory requirements. Consequently, as compliance costs rise, there is a risk that innovation may be stifled and the development of groundbreaking AI technologies within the EU could grind to a halt.[93]

87 European Parliament resolution on the urgent need to revise the Medical Devices Regulation (2024/2849(RSP)), available at: https://www.europarl.europa.eu/doceo/document/RC-10-2024-0123_EN.html .

88 *Supra* note 22.

89 Ellen O'Regan, Europe's GDPR privacy law is headed for red tape bonfire within 'weeks', Politico (3 April 2025), available at: https://www.politico.eu/article/eu-gdpr-privacy-law-europe-president-ursula-von-der-leyen/.

90 Pieter Haeck, EU opens door to reworking AI rulebook, Politico (9 April 2025), https://www.politico.eu/article/how-eu-did-full-180-artificial-intelligence-rules/; see also European Commission, Commission sets course for Europe's AI leadership with an ambitious AI Continent Action Plan, press release (9 April 2025), available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1013

91 European Commission, Directorate General for Communication, Commission proposes to cut red tape and simplify business environment (26 February 2025), available at: https://commission.europa.eu/news/commission-proposes-cut-red-tape-and-simplify-business-environment-2025-02-26_en .

92 See e.g. Jonas Tallberg, Magnus Lundgren, and Johannes Geith, *AI Regulation in the European Union: Examining Non-State Actor Preferences*, Business and Politics 26, no. 2 (2024): 241–267, https://doi.org/10.1017/bap.2024.10 .

93 See Aboy, Vayena, Minssen, *supra* note 8.

Moreover, the potential for regulatory divergence between the EU and other regions—such as the United States or China—could further undermine the EU's competitive position. As AI technologies continue to advance at an unprecedented pace, the more flexible regulatory environments in other jurisdictions may attract investments and talent away from Europe. Companies may choose to relocate their AI research and development activities to regions with fewer restrictions, undermining the EU's position as a global leader in AI innovation. This is particularly concerning given that the AI sector is characterized by rapid iterations and the need for agile responses to market demands.

On the other hand, the AIA also has the potential to enhance the EU's competitive edge in certain areas. By establishing clear ethical guidelines and standards, the Act could foster trust among consumers and businesses in AI technologies. In a market increasingly concerned about privacy, security, and the ethical implications of emerging technologies, the EU's proactive stance on regulating AI may serve as a differentiating factor. European companies that comply with these regulations can position themselves as leaders in responsible AI, appealing to consumers who prioritize ethical considerations. This trust can translate into market advantages, particularly in sectors like healthcare, finance, and autonomous systems, where ethical concerns are paramount.

Furthermore, the AIA emphasizes the importance of human oversight and accountability in AI systems. This focus on responsible AI practices could drive innovation in developing transparent and explainable AI models. Such advancements may not only benefit European companies but could also set a global standard for AI practices. If the EU can successfully cultivate a reputation for high-quality, ethical AI, it may attract international collaborations and partnerships, reinforcing its position in the global AI ecosystem.

In conclusion, the AIA presents a double-edged sword for the technological competitiveness of the EU. While the potential for regulatory burdens may hinder innovation and drive talent elsewhere, the emphasis on ethical standards and consumer trust could establish Europe as a leader in responsible AI. Striking the right balance between regulation and innovation will be crucial in determining whether the EU can maintain its technological competitiveness in an increasingly AI-driven world. As global dynamics shift, ongoing dialogue among policymakers, industry stakeholders, and the public will be essential to ensure that the AIA serves as a catalyst for growth rather than a barrier to progress.

II.IV LESSONS FOR OTHER COUNTRIES

Countries around the world can learn several important lessons from the EU's AIA as they navigate the complexities of AI regulation and development. One of the primary lessons is the necessity of finding a balance between regulatory oversight and fostering innovation. While it is critical to establish safety and ethical standards for AI technologies, overly burdensome regulations can stifle creativity and deter investment. Countries should aim to create flexible regulatory frameworks that can adapt to the fast-paced nature of AI development, allowing for innovation while ensuring public safety and upholding ethical principles.

The EU's emphasis on ethical AI and transparency highlights the importance of building public trust in AI systems. Countries should prioritize developing guidelines that promote responsible AI use, which includes ensuring that AI applications are explainable and accountable. Establishing clear ethical standards can not only enhance consumer confidence but can also position nations as leaders in responsible AI, attracting businesses and talent seeking to operate in a trustworthy environment.

The AIA's development illustrates the value of engaging a broad range of stakeholders, including industry leaders, academics, and civil society. Inclusive dialogue can ensure that regulations are grounded in practical realities and address the concerns of all parties affected. Countries should adopt collaborative approaches to policymaking, ensuring that regulations reflect diverse perspectives and foster a healthy ecosystem for AI innovation.

As AI technologies transcend borders, international cooperation is essential. Countries should seek to harmonize regulations and share best practices to avoid fragmentation. Collaborative efforts can enhance global standards, encourage knowledge sharing, and facilitate cross-border investments in AI research and development.

Finally, investing in regulatory capacities, education, and workforce training is crucial. As AI continues to evolve, countries must equip their authorities and populations with the skills needed to thrive in an AI-driven economy. This proactive approach will ensure that nations remain competitive and can adapt to the changing technological landscape. This also means that sufficient funds will have to be directed to the authorities, named bodies and stakeholders that are supposed to monitor, support, and enforce complex regulatory set-ups. As other countries consider how the EU's approach to AI governance pans out, part of this evaluation must consider if its complexity is nevertheless deftly handled by a set of well-resourced and highly competent actors. If the actors and mechanisms involved in the monitoring, supporting, and enforcement of the AIA fall short

of expectations, other countries might prefer more streamlined and simplified regimes. It will be for other countries to decide whether a mismatch between the regulatory burden and institutional capacity, in the hopes this would be a teething problem during the Regulation's infancy, would be a preferable approach to having fewer rules that are easier to comply with and more likely to be enforced.

By learning from the EU's experience, other countries can strategically navigate the challenges and opportunities presented by AI, fostering a balanced, ethical, and innovative environment with a feasible and enforceable regulatory set-up.

III. CONCLUSIONS

Any number of criticisms should not detract from the fact that the EU's AIA is a groundbreaking legislation that will profoundly shape the future of medical AI, both within and potentially beyond Europe. Its role within the EU's regulatory ecosystem is multifaceted as it interacts with existing legislation, promotes ethical AI development, and encourages international cooperation. Its extraterritorial reach, the size of the EU market, and the benefits of harmonized global best practices, make it difficult for foreign stakeholders to ignore its implications. Whilst the likelihood of it being used as a template for other jurisdictions to copy near verbatim seems to be increasingly small, there are valuable lessons for other jurisdictions concerning the development and implementation of comprehensive AI regulation.

The drafting of the AIA demonstrates that the universally accepted desirability of promoting innovation and protecting individual rights does not create a consensus of how to achieve these aims. The AIA, as a pioneering regulatory framework, offers valuable lessons for other countries navigating the complexities of AI governance. By learning from both the advances and challenges encountered in the EU's approach, jurisdictions around the world can develop more effective, ethical, and adaptable regulatory frameworks that promote innovation while safeguarding public interests.

Being the first to regulate does not necessarily mean the EU created the best AI regulation possible, but how actors cope with these rules will mean the AIA will continue to play a crucial role in shaping the future of AI governance in Europe and beyond. By prioritizing flexibility, support for research, and workforce development, countries can position themselves to harness the transformative potential of AI while ensuring that its deployment aligns with societal values,

legal framework and ethical considerations. Ultimately, the goal should be to foster an environment that encourages innovation while safeguarding public interests, thereby ensuring a competitive edge in the rapidly evolving global AI landscape.

For the time being, foreign stakeholders that want to invest in European markets, including companies from Latin America, should be proactive and stay informed about the AIA's implications for AI systems. To successfully navigate the AIA's impact and to avoid harsh fines, they should carefully assess their obligations, invest in compliance measures, and engage with EU regulators during its implementation phases. Even after the AIA fully enters into force, the European Commission may adopt implementing acts and delegated acts that may impact key provisions.[94] The regime will thus continue to evolve through interpretation, application, guidelines, clarifying exceptions, and any potential amendments. Stakeholders, including Latin American companies, should actively provide input to the development of adaptive governance frameworks, i.e., not only to tackle the challenges and seize the opportunities presented by the AIA while maintaining access to the EU market, but also to contribute to developing their own domestic and regional approaches for the governance of AI.

Last, but not least, there remains an urgent need to establish international treaties and agreements on the most fundamental guardrails that should apply to the governance and deployment of AI.[95] While national approaches might differ with regard to regulatory details that make a globally aligned approach difficult, this does not make the need for agreement on international obligations to protect and promote fundamental values any less important nor any less pressing. The need for such basic "guardrails" is more urgent than ever when considering the enormous advances, opportunities and risks of dual use scenarios with other emerging technologies, such as neurotechnologies, synthetic biotechnologies, quantum technologies, and the combination of these with AI.

Acknowledgements: The authors would like to thank Emma Kondrup for her assistance with this chapter, both editorial and in developing the illustrations for this manuscript. Any use of generative AI in this manuscript adheres to ethical guidelines for use and acknowledgment of generative AI in academic research, as outlined in Porsdam Mann, S., Vazirani, A.A., Aboy, M.; Minssen, T. *et al.*

94 *Supra* note 70.

95 See also Urs Gasser and Viktor Mayer Schönberger. *Guardrails: Guiding Human Decisions in the Age of AI.* Princeton University Press, 2024. https://doi.org/10.2307/jj.6605394.

Guidelines for ethical use and acknowledgement of large language models in academic writing. *Nat Mach Intell* 6, 1272–1274 (2024). https://doi.org/10.1038/s42256-024-00922-7. Each author has made a substantial contribution to the work, which has been thoroughly vetted for accuracy, and assumes responsibility for the integrity of their contributions. T.M.; H.L.S. & S.P.M. were supported in part by a Novo Nordisk Foundation grant for a scientifically independent International Collaborative Bioscience Innovation & Law Programme (Inter-CeBIL Programme, grant NNF23SA0087056). T.M. was further funded in part by the European Union Classica project (grant 101057321). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency.

BIBLIOGRAPHY

Aboy, M., Minssen, T. & Vayena, E. *Navigating the EU AIA: implications for regulated digital medical products. npj Digit. Med.* 7, 237 (2024). https://doi.org/10.1038/s41746-024-01232-3.

Almada, Marco and Radu, Anca *The Brussels Side-Effect: How the AIA Can Reduce the Global Reach of EU Policy. German Law Journal* (2024).

Almada, Marco and Petit, Nicolas, *The EU AIA: A Medley of Product Safety and Fundamental Rights? Robert Schuman Centre for Advanced Studies Research Paper* No. 2023/59 (October 18, 2023).

Applied AI Initiative, *AIA Impact Survey: Pan-European Survey of AI Startups and VCs about the Impact of the AIA*, AppliedAI.De, (December 12, 2022). https://www.appliedai.de/assets/files/AI-Act-Impact-Survey_Slides_Dec12.2022.pdf.

Biden Jr., Joseph R., *Executive Order 14141: Advancing United States Leadership in Artificial Intelligence Infrastructure,* Federal Register 90, no. 5469 (January 17, 2025): 5469–5489.

Bradford Anu, *The Brussels Effect: How the European Union Rules the World*, New York: Oxford University Press (2020). https://doi.org/10.1093/oso/9780190088583.001.0001.

Buocz, Thomas, Pfotenhauer, Sebastian, and Eisenberger, Iris, *Regulatory Sandboxes in the AIA: Reconciling Innovation and Safety? Law, Innovation and Technology* 15, no. 2 (2023): 357–389. https://doi.org/10.1080/17579961.2023.2245678.

Cantero Gamito, Marta, and Marsden, Christopher T. *AI Co-Regulation? The Role of Standards in the EU AIA*. International Journal of Law and Information Technology 32 (2024). https://doi.org/10.1093/ijlit/eaae011.

Case C-184/20, OT v. Vyriausioji tarnybinės etikos komisija, ECLI:EU:C:2022:601, August 1, 2022.

Colonna, Liane, *The AIA's Research Exemption: A Mechanism for Regulatory Arbitrage?* in *YSEC Yearbook of Socio-Economic Constitutions 2023: Law and the Governance of AI*, ed. Andreas Moberg and Eduardo Gill-Pedro (Cham: Springer Nature Switzerland, 2024).

Crawford, Gail E. *et al.*, *EU AI Office Seeks Stakeholder Engagement in Drafting AI Act Implementing Rules*, Latham & Watkins LLP (July 31, 2024). https://www.lw.com/en/insights/eu-ai-office-seeks-stakeholder-engagement-in-drafting-ai-act-implementing-rules.

Creative Commons *et al.*, *Supporting Open Source and Open Science in the EU AIA* (July 2023).

Draghi, Mario, The future of European competitiveness, available at: https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en (EU Commission September 2024)

Ebers, Martin, *Truly Risk-Based Regulation of AI: How to Implement the EU's AIA*, *European Union Law Working Papers* (October 21, 2024), Stanford-Vienna Transatlantic Technology Law Forum.

European Commission. *Commission Guidelines on Prohibited AI Practices Established by Regulation (EU) 2024/1689 (AIA).* C(2025) 884 final, Brussels (February 4, 2025), https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

European Commission, *Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises* (Text with EEA relevance) (notified under document number C(2003) 1422) OJ L 124, (20 May 2003):36-41.

European Commission, *Commission Work Programme 2025*, last modified February 12, 2025, https://commission.europa.eu/strategy-and-policy/strategy-documents/commission-work-programme/commission-work-programme-2025_en.

European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Building Trust in Human-Centric AI*, COM(2019) 168 final (April 8, 2019).

European Commission, *Directorate-General for Communications Networks, Content and Technology, Ethics guidelines for trustworthy AI,* Publications Office (2019). https://data.europa.eu/doi/10.2759/346720.

European Commission, *First Rules of the AI Act Are Now Applicable*, *Shaping Europe's Digital Future* (February 3, 2025), https://digital-strategy.ec.europa.eu/en/news/first-rules-artificial-intelligence-act-are-now-applicable.

European Commission, *Guidelines on the Definition of an AI System Established by the AIA* (2024), https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application.

European Commission, *High-Level Group on the Digital Markets Act: Public Statement on Artificial Intelligence*, Digital Markets Act (May 22, 2024). https://digital-markets-act.ec.europa.eu/high-level-group-digital-markets-act-public-statement-artificial-intelligence-2024-05-22_en

European Commission. *Liability Rules for Artificial Intelligence.* Last modified February 12, 2025. https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en.

European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on AI (AI Act) and Amending Certain Union Legislative Acts*, COM(2021) 206 final (April 21, 2021).

European Commission, A Competitiveness Compass for the EU, Brussels, 29.1.2025 COM(2025) 30 final, available at: https://commission.europa.eu/document/download/10017eb1-4722-4333-add2-e0ed18105a34_en .

European Commission, Commission sets course for Europe's AI leadership with an ambitious AI Continent Action Plan, press release (9 April 2025), available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1013 .

European Commission, Directorate General for Communication, Commission proposes to cut red tape and simplify business environment (26 February 2025), available at: https://commission.europa.eu/news/commission-proposes-cut-red-tape-and-simplify-business-environment-2025-02-26_en .

European Data Protection Supervisor. *Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the Light of Legislative Developments* (October 23, 2023):11. https://www.edps.europa.eu/system/files/2023-10/2023-0137_d3269_opinion_en.pdf

European Parliament and Council. *Regulation (EU) No 536/2014 on Clinical Trials on Medicinal Products for Human Use and Repealing Directive 2001/20/EC* (Text with EEA relevance), OJ L 158, May 27, 2014, 1–76.

European Parliament and Council. *Directive (EU) 2024/2853 on Liability for Defective Products and Repealing Council Directive 85/374/EEC* (Text with EEA relevance), PE/7/2024/REV/1, OJ L 2024/2853, (November 18, 2024).

European Parliament and Council. *Regulation (EU) 2023/988 on General Product Safety, Amending Regulation (EU) No 1025/2012 and Directive (EU) 2020/1828, and Repealing Directive 2001/95/EC and Council Directive 87/357/EEC* (Text with EEA relevance), PE/79/2022/REV/1, OJ L 135, May 23, 2023, 1–51.

European Parliament and Council. *Regulation (EU) 2022/1925 on Contestable and Fair Markets in the Digital Sector and Amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)* (Text with EEA relevance), PE/17/2022/REV/1, OJ L 265, October 12, 2022, 1–66.

European Parliament and Council. *Regulation (EU) 2022/2065 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act)* (Text with EEA relevance), PE/30/2022/REV/1, OJ L 277, October 27, 2022, 1–102.

European Parliament and Council. *Regulation (EU) 2017/746 on In Vitro Diagnostic Medical Devices and Repealing Directive 98/79/EC and Commission Decision 2010/227/EU* (Text with EEA relevance), OJ L 117, May 5, 2017, 176–332.

European Parliament and Council. *Regulation (EU) 2017/745 on Medical Devices, Amending Directive 2001/83/EC, Regulation (EC) No 178/2002, and Regulation (EC) No 1223/2009, and Repealing Council Directives 90/385/EEC and 93/42/EEC* (Text with EEA relevance), OJ L 117, May 5, 2017, 1–175.

European Parliamentary Research Service, *Algorithmic discrimination under the AI Act and the GDPR* (February, 2025). https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2025)769509.

European Union, *Charter of Fundamental Rights of the European Union*, *Official Journal of the European Communities,* C 364/1 (December 18, 2000).

European Union, *Consolidated Version of the Treaty on the Functioning of the European Union (TFEU) Official Journal of the European Union* C 115 (May 9, 2008):94–95.

European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council on AI (AIA)*, *Official Journal of the European Union,* L 311/1 (July 12, 2024).

Farrell, Jimmy and Shamir, Afek, *An Introduction to Codes of Practice for the AI Act,* EU Artificial Intelligence Act, last modified October 30, 2024, https://artificialintelligenceact.eu/introduction-to-codes-of-practice/.

Gasser, Urs, and Mayer-Schönberger, Viktor. *Guardrails: Guiding Human Decisions in the Age of AI*. Princeton University Press, 2024. https://doi.org/10.2307/jj.6605394.

Gornet, Mélanie, and Maxwell, Winston. 2024. *The European Approach to Regulating AI through Technical Standards*. Internet Policy Review **13** (3), https://policyreview.info/articles/analysis/regulating-ai-through-technical-standards.

Grozdanovski, Ljupcho and De Cooman, Jérôme, *On the Obsolescence of Empirical Knowledge in the Age of AI, Rutgers Computer & Technology Law Journal* 49, no. 2 (2023): 123–145.

Haataja, Meeri and Joanna J. Bryson, *What Costs Should We Expect from the EU's AIA?*, *SocArXiv* (2021).

Haeck, Pieter EU opens door to reworking AI rulebook, Politico (9 April 2025), https://www.politico.eu/article/how-eu-did-full-180-artificial-intelligence-rules/.

Ho-Dac, Marion, *Considering Fundamental Rights in the European Standardisation of AI: Nonsense or Strategic Alliance?*, *arXiv* (January 23, 2024).

Holst, Luca **et al.**, *The Impact of the EU AIA's Transparency Requirements on AI Innovation*, in *Proceedings of the 19th International Conference on Wirtschaftsinformatik*, Würzburg, Germany (2024).

International Association of Privacy Professionals, *EU AI Act Stakeholder Map*. IAPP (May 2024). https://iapp.org/resources/article/eu-ai-act-stakeholder-map/.

Jones, Joe, *EU AIA: Next Steps for Implementation*, *International Association of Privacy Professionals* (January 2025), https://iapp.org/resources/article/eu-ai-act-timeline/.

Kantner, Cathleen, *What Is a European Identity? The Emergence of a Shared Ethical Self-Understanding in the European Union*, *EUI RSCAS Working Papers* 2006/28 (2006).

Kasirzadeh, Atoosa, *Two Types of AI Existential Risk: Decisive and Accumulative*. arXiv preprint arXiv:2401.07836 (2024).

van Kolfschooten, Hannah, *EU Regulation of AI: Challenges for Patients' Rights,* Common Market Law Review 59, no. 1 (2022): 81–112.

van Kolfschooten, Hannah, and Shachar, Carmel, *The Council of Europe's AI Convention (2023–2024): Promises and Pitfalls for Health Protection,* Health Policy 138 (December 2023): 104935, https://doi.org/10.1016/j.healthpol.2023.104935.

Mann, Sebastian Porsdam, Cohen, I. Glenn and Minssen, Timo, *The EU AIA: Implications for U.S. Health Care,* NEJM AI 1, no. 11 (October 2024), https://ai.nejm.org/doi/pdf/10.1056/AIp2400449.

Marsden, Christopher T. and Christou, George *Artificially Intelligent Regulation: Global Norms, International Political Economy and the Brussels Effect*, in *Proceedings of the 2024 International Conference on AI and Regulation* (2024).

Ministry of Foreign Affairs of Japan, *Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems* (October 30, 2023).

Moille, Céline, Popescu, Adelina and Scott, Gretchen, *EU AIA Implementation Timeline*, *Goodwin Insights* (October 10, 2024), https://www.goodwinlaw.com/en/insights/publications/2024/10/insights-technology-aiml-eu-ai-act-implementation-timeline.

Neergaard, Ulla and de Vries, Sybe, *The Interaction between Free Movement Law and Fundamental Rights in the (Digital) Internal Market,* Renforce Working Paper 1-2023 (2023): 1–30.

O'Regan, Ellen, Europe's GDPR privacy law is headed for red tape bonfire within 'weeks', Politico (3 April 2025), available at: https://www.politico.eu/article/eu-gdpr-privacy-law-europe-president-ursula-von-der-leyen/.

Organisation for Economic Co-operation and Development, *OECD Framework for the Classification of AI Systems*, Paris: *OECD Publishing* (2022).

Orwat, Carsten *et al.*, *Normative Challenges of Risk Regulation of AI*, NanoEthics 18, no. 2 (2024): 11, https://doi.org/10.1007/s11569-024-00454-9.

Overseas Development Institute. *International AI Governance Must Be Truly Global*, ODI (October 25, 2023). https://odi.org/en/insights/international-ai-governance-must-be-truly-global/.

Pelkmans, Jacques and Renda, Andrea *Does EU Regulation Hinder or Stimulate Innovation?*, *CEPS Special Report No. 96* (November 19, 2014), Brussels: *Centre for European Policy Studies*.

Podszun, Rupprecht, *From Competition Law to Platform Regulation: Regulatory Choices for the Digital Markets Act,* Economics 17, no. 1 (2023): 20220037, https://doi.org/10.1515/econ-2022-0037.

Price II, W. Nicholson and Cohen, I. Glenn, *Locating Liability for Medical AI,* DePaul Law Review 73, no. 2 (2024): 339–368.

PYMNTS, *European Commission Withdraws Draft Rules on Technology Patents, AI Liability and Consumer Privacy* (February 12, 2025), https://www.pymnts.com/news/regulation/2025/european-commission-withdraws-draft-rules-on-technology-patents-ai-liability-and-consumer-privacy/.

Shao, Shuai *et al.*, *Environmental Regulation and Enterprise Innovation: A Review,* Business Strategy and the Environment 29, no. 3 (2020): 1465–1478, https://doi.org/10.1002/bse.2446.

Siegmann, Charlotte and Anderljung, Markus, *The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market*, Centre for the Governance of AI (Aug. 2022). https://cdn.governance.ai/Brussels_Effect_GovAI.pdf

Tallberg, Jonas, Lundgren, Magnus and Geith, Johannes, *AI Regulation in the European Union: Examining Non-State Actor Preferences*, Business and Politics 26, no. 2 (2024): 241–267, https://doi.org/10.1017/bap.2024.10.

Teo, Sue Anne, *How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework*, Nordic Journal of Human Rights 40, no. 1 (2022): 216–34, https://doi.org/10.1080/18918131.2022.2073078.

The White House. *Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.* Statements and Releases (October 30, 2023). https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

Trump, Donald J., *Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence*, Federal Register 90, no. 5469 (January 23, 2025): 5469–5471. https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence.

Veale, Michael and Borgesius, Frederik Zuiderveen, *Demystifying the Draft EU AI Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach,* Computer Law Review International 22, no. 4 (2021): 97–112, https://doi.org/10.9785/cri-2021-220402.

Wachter, Sandra*, Limitations and Loopholes in the EU AIA and AI Liability Directives: What This Means for the European Union, the United States, and Beyond,* Yale Journal of Law & Technology 26, no. 3 (2024):685.

EL FUTURO DEL DERECHO DE AUTOR EN AMÉRICA LATINA EN LA ERA DE LA INTELIGENCIA ARTIFICIAL

SUMARIO

I. Introducción **II.** El derecho de autor en América Latina. Antecedentes y características **III.** Preocupaciones y preguntas legales en torno a la IA y el derecho de autor **IV.** Desafíos y limitaciones actuales del derecho de autor ante el avance de la IA generativa **V.** El futuro del derecho de autor en la era de la IA **VI.** Proyectos de regulación en la región **VII.** Conclusión **VIII.** Bibliografía.

> MARÍA DE LOURDES VÁZQUEZ

Decana de la Escuela de Derecho de la Universidad de San Andrés (UdeSA, Buenos Aires, Argentina). Directora de Maestría de la Propiedad Intelectual e Innovación organizada en conjunto con la Organización Mundial de la Propiedad Intelectual (OMPI) y UdeSA. Magister (LL.M.) por Harvard Law School (Cambridge, Massachusetts)

> MALENA MANCINI

Abogada por la Universidad de Buenos Aires (UBA), especializada en propiedad intelectual y derecho del entretenimiento. Magister en Propiedad Intelectual e Innovación por la Universidad de San Andrés. Docente en la Facultad de Derecho de la UBA, la Universidad de Ciencias Empresariales y Sociales, la Universidad del Museo Social Argentino y la Universidad de San Andrés.

I. INTRODUCCIÓN

Al igual que en el resto del mundo, en América Latina estamos experimentando el avance vertiginoso de la inteligencia artificial (IA). En los últimos años la IA se ha vuelto fundamental en múltiples sectores y su uso ha aumentado de manera exponencial. Su acelerado crecimiento ha abierto una serie de debates respecto a la legalidad y moralidad de su desarrollo e implementación, entre ellos se destacan importantes discusiones en torno a la propiedad intelectual en general y al derecho de autor en particular. En este trabajo abordaremos los desafíos éticos y legales que plantea esta tecnología disruptiva, enfocándonos en las problemáticas centrales que surgen a raíz de la implementación de los sistemas de IA, en particular, aquellos de IA generativa.

Los países de la región se encuentran en diversas etapas de adaptación de sus marcos regulatorios para abordar estas nuevas realidades. Aunque aún se están explorando alternativas, algunos países han comenzado a considerar cómo ajustar sus leyes para equilibrar la protección de los autores humanos con la innovación tecnológica que permite la IA. Este es un momento crucial para la región, ya que lo que se decida hoy impactará el acceso al conocimiento, la preservación de la cultura y la competitividad digital en el futuro. En este contexto, América Latina se enfrenta a la oportunidad de ser un actor relevante en la discusión global, buscando desarrollar un enfoque que fomente tanto la innovación como la justicia en la protección de los derechos de autor.

La IA es un fenómeno global que ha captado tanto el entusiasmo como la preocupación de la comunidad internacional. Los avances en esta tecnología cruzan fronteras, afectando tanto a países desarrollados como a naciones en desarrollo. Hoy en día, los modelos de IA se entrenan con grandes cantidades

de datos recopilados a nivel mundial, a menudo mediante técnicas de *'scraping'* de contenidos disponible en línea, las cuales captan entre otras cosas material protegido por derechos de autor. Esto genera un impacto directo en autores y creadores de todas las regiones, planteando serias preguntas sobre cómo proteger sus derechos de forma efectiva y justa.

Frente a este contexto, surge la pregunta de si es necesario y posible establecer un marco regulatorio global que contemple los derechos de autores y creadores en todo el mundo. Dado lo ambicioso de este objetivo, una alternativa más alcanzable para América Latina podría ser la cooperación regional. La armonización de legislaciones entre los países andinos, Mercosur y otras alianzas regionales podría ser un primer paso, fortaleciendo un enfoque común que atienda tanto la innovación como la protección de derechos. La colaboración regional permitiría establecer principios básicos que orienten la regulación de la IA en nuestros países y nos preparen para futuros acuerdos internacionales en la materia.

En lo relativo a derecho de autor, las cuestiones legales relacionadas con la IA se suelen dividir principalmente en dos áreas: las problemáticas que surgen del *input* (los datos utilizados para entrenar los algoritmos) y las cuestiones relativas a la autoría sobre el *output* (los resultados generados por los sistemas de IA). En cuanto al *input*, surgen preguntas sobre si es necesario obtener autorización de los titulares de derechos para usar sus obras en el entrenamiento de algoritmos, se cuestiona si los autores deberían ser compensados por ello y si ese uso podría considerarse *fair use* bajo la ley anglosajona. En el ámbito del *output*, la preocupación gira en torno a si los resultados generados por la IA pueden considerarse obras originales o derivadas, si deben ser protegidos por derechos de autor y, en caso afirmativo, quién sería el titular de esos derechos.

El desarrollo de la IA generativa en los últimos años ha dejado en evidencia sus capacidades, lo cual ha despertado preguntas sobre cuestiones éticas y legales fundamentales en el ámbito del derecho de autor, como así también preocupaciones respecto al impacto de esta tecnología en el mercado creativo. Conceptos como creatividad artificial están comenzando a cobrar relevancia, llevando a reevaluar las bases y objetivos del derecho de autor. A su vez, una ola de conflictos legales que tiene en el centro a empresas desarrolladoras de modelos de IA dejan en evidencia las deficiencias de la legislación actual para dar respuestas a las necesidades de esta tecnología.

El futuro del derecho de autor en la era de la IA está siendo debatido en todo el mundo. La rápida evolución legal ha provocado lo que se conoce como las "guerras del copyright," un debate continuo sobre el rol del derecho

de autor en la preservación de la cultura y el conocimiento.[1] En varios países de América Latina se observan proyectos de ley relacionados con IA, entre los cuales algunos incluso tratan cuestiones relacionadas al derecho de autor, mas no hay consenso sobre cómo abordar estos temas. Las numerosas discusiones doctrinarias y conflictos legales que han surgido en los últimos años dejan en evidencia que la legislación vigente no está preparada para hacer frente a las exigencias del nuevo paradigma tecnológico y creativo que ha planteado la IA generativa.

Existen distintos posibles caminos para el derecho de autor en América Latina, sin embargo, para poder hablar del futuro es necesario entender el presente. Este trabajo comienza con una breve descripción de qué son los derechos de autor y cuál es la situación regulatoria actual en la materia en nuestra región. Considerando que gran parte de la discusión está ocurriendo en Estados Unidos, es importante a su vez remarcar las diferencias entre el sistema de Copyright anglosajón y el derecho de autor continental que hemos adoptado en Latinoamérica, ya que estas pueden dar lugar a un tratamiento ampliamente distinto de estos temas. Expuesto lo anterior, quedarán de manifiesto las preocupaciones y preguntas legales que surgen en torno a la IA y el derecho de autor, lo cual permitirá abordar la nueva realidad que plantea el uso de estas herramientas en las industrias creativas y los conflictos que han surgido a raíz de ello. Resulta esencial para el tema que se aborda en el presente analizar los principales casos judiciales que se están disputando en materia de derecho de autor e IA generativa, siendo que su resolución podría clarificar algunas de las cuestiones aquí tratadas.

Una vez planteado el panorama completo, se analizará de manera crítica la regulación vigente en materia de derecho de autor ante el avance de la IA generativa, detallando cuáles son los desafíos y limitaciones que esta presenta. Expuestas las deficiencias de los sistemas vigentes, se identificarán tres posibles escenarios para el futuro de la regulación de las creaciones en la era de la IA: actualizar del derecho de autor, buscar respuestas en otras ramas del derecho o plantear un nuevo marco normativo *sui generis* que contemple las particularidades de la situación. Para finalizar, se hará mención a proyectos de ley en la región que abarcan estos temas aportando un panorama de los potenciales avances que se avecinan.

1 Lessig, L, Cultura libre: Cómo los grandes medios usan la tecnología y las leyes para encerrar la cultura y controlar la creatividad, LOM Ediciones, Primera edición en Chile, 2005. I.S.B.N: 956-282-745-3. p.148 (Título Original: “Free Culture: How Big Media uses Technology and the Law to lock down Culture and control Creativity”)

II. EL DERECHO DE AUTOR EN AMÉRICA LATINA. ANTECEDENTES Y CARACTERÍSTICAS

Antes de analizar los desafíos que presenta la IA en relación con el derecho de autor y analizar los cambios que estos impulsan, es preciso entender de qué se trata esta rama del derecho que hoy aparece tan entrelazada a las discusiones legales que rodean a la tecnología del momento. Para ello no basta con definir qué son los derechos de autor, sino que es necesario hacer una breve mención a su origen a fin de lograr comprender de dónde surgieron estos derechos, cuál fue la necesidad que los motivó y cuál ha sido el fundamento económico detrás de su regulación.

En resumidas cuentas, el derecho de autor se refiere a los derechos de los autores y creadores sobre sus obras.[2] En gran parte del mundo, la existencia de estos derechos no requiere más formalidad que su expresión, es decir, que no depende de un registro ni documento alguno. Este principio fue receptado en uno de los tratados internacionales más importantes en materia de derecho de autor, el Convenio de Berna para la Protección de las Obras Literarias y Artísticas (1886) del cual hoy forman parte 181 países.

Si bien desde la antigua Grecia se observa un reconocimiento del vínculo entre un autor y su obra, no se habla de derechos de propiedad intelectual en torno a estas sino hasta mediados del siglo XV con la invención de la imprenta. Hasta ese momento, las copias de libros se realizaban de manera manual, lo cual representaba un proceso laborioso, sujeto a errores y propenso a censura. La reproducción mecánica permitió producir una gran cantidad de copias exactas de una misma obra, reduciendo los tiempos y costo del proceso, aumentando a su vez la disponibilidad y el acceso al material impreso. Si bien el mercado literario en esa época no era lo que es hoy, esta nueva tecnología presentó una oportunidad para la difusión masiva de conocimiento e ideas, incluso aquellas ideas disidentes y con críticas hacia la Corona y la Iglesia. Es en este contexto en donde surgen las primeras normas que regulan la reproducción de obras. Estas no otorgaban derechos a los autores, sino que más bien se trataban de permisos o licencias concedidas a impresores, editores o gremios de estos, que habilitaban la impresión de ciertos libros de manera exclusiva.[3] Este sistema se

2 "Derecho de Autor", WIPO - World Intellectual Property Organization, s. f., https://www.wipo.int/copyright/es/.

3 Camilo Mirosevic Verdugo, "Origen y Evolución del Derecho de Autor Con Especial Referencia Al Derecho Chileno", *Revista de Derecho de La Pontificia Universidad Católica de Valparaíso* XXVIII (2007): 35-82, https://www.redalyc.org/articulo.oa?id=173620169001.

basaba en "privilegios" o monopolios que ponían el foco en la reproducción y difusión de las obras, dejando de lado los derechos de los autores.[4]

LA PRIMERA LEY DE COPYRIGHT

Esto cambia cuando en 1710 surge en Inglaterra el llamado "Estatuto de la Reina Ana",[5] definido como *"An Act for the Encouragement of Learning…"*[6] –una ley para fomentar el aprendizaje–, resaltando un propósito utilitario al señalar el beneficio social de la regulación del derecho de autor. Esta norma introduce un cambio significativo en la concepción de la protección de los derechos literarios, ya que la protección se centra en el autor como titular de sus derechos intelectuales, estableciendo además el principio de temporalidad en la vigencia de dichos derechos.[7]

Desde el siglo XVIII con la implementación de esta ley, las normativas que regulan los derechos de los autores y creadores han replicado esta estructura reconociendo la titularidad de la obra en cabeza del autor y otorgándole a este derechos exclusivos durante un período de tiempo determinado para su explotación. Uno de los primeros en adoptar estos principios fue Estados Unidos, inicialmente en su Constitución y luego en la *U.S. Copyright Law.*

DOS TRADICIONES: COPYRIGHT ANGLOSAJÓN Y DERECHO DE AUTOR CONTINENTAL

Las regulaciones en materia de derecho de autor que se fueron adoptando desde entonces en las distintas jurisdicciones del mundo reflejan la existencia de dos corrientes marcadas. Una es el sistema de Copyright anglosajón que rige en Estados Unidos, Reino Unido y los países del Commonwealth. Este entiende que el derecho de autor tiene una justificación utilitarista, otorgando derechos

4 Christian Schmitz Vaccaro, "Evolución de la Regulación Internacional de la Propiedad Intelectual," Revista la Propiedad Inmaterial 17 (2013): 63-92.

5 El título oficial de la norma era "Ley de fomento del aprendizaje por la que se otorga el derecho sobre las copias de libros impresos a los autores o compradores de las copias, durante el plazo en ella establecido" y fue promulgada por el Parlamento británico en 1710.

6 An Act for the Encouragement of Learning, from Suarez, M. F., & Woudhuysen, H. R. (2010). *The oxford companion to the book*. Oxford University Press.

7 Catherine Jewell, "En defensa del derecho de autor: visión de los interesados", WIPO - World Intellectual Property Organization, abril de 2014, https://www.wipo.int/wipo_magazine/es/2014/02/article_0004.html.

exclusivos a los titulares para que exploten sus obras ya que esto presenta un beneficio para la sociedad, la cultura y el arte.[8] La otra es la conocida como 'Sistema Continental' y sienta la base en el derecho natural de quien crea,[9] dando lugar a un derecho que nace del propio acto de creación y que es inherente al autor. Esta acepción del derecho de autor se caracteriza por poseer una dimensión moral que reconoce una serie de derechos al autor que son imprescriptibles, intransferibles e irrenunciables. En ambos casos se reconoce al autor derechos exclusivos sobre la obra, siendo esta exclusividad la que le permite obtener un rédito económico de su trabajo.

En nuestra región se ha adoptado la concepción continental del derecho de autor que deriva de la tradición civilista que rige en la mayor parte de América Latina. Sin embargo, no todas las legislaciones han seguido el mismo modelo en la regulación de la materia. Algunas poseen normas más extensas y detalladas, mientras otras aún cuentan con articulados desactualizados e incompletos. Sin perjuicio de las diferencias, y de cómo cada jurisdicción decida abordar el tema, no suele haber mayores diferencias en la protección que prevén.

Cada uno de estos sistemas, el sistema de Copyright y el de derecho de autor, posee sus particularidades y algunas diferencias marcadas. Además de las ya mencionadas respecto a sus fundamentos y los derechos que reconocen a los autores, estos sistemas prevén distintas estructuras de excepciones. Las excepciones son limitaciones al derecho exclusivo del titular de la obra que se imponen por ley, en la práctica representan situaciones en las cuales las obras pueden ser utilizadas sin autorización. Mientras el derecho de autor aplica un enfoque restringido en lo que respecta a las excepciones que admite, previendo en su cuerpo legal un sistema "cerrado" de excepciones[10] –una cantidad acotada de escenarios en los cuales no es necesario procurar el consentimiento del titular para usar una obra–, el sistema de Copyright adopta un enfoque amplio que permite el uso de una obra sin autorización bajo ciertas circunstancias las cuales

8 Noemí Pulido Pavón, Luis Palma Martos, y Luis F. Aguado, "Derechos de Autor. Enfoque Económico, Evolución y Perspectivas", Revista de Economía Institucional 18, n.o 35 (29 de noviembre de 2016): 151, https://doi.org/10.18601/01245996.v18n35.08.

9 Pavón, Martos, y Aguado (2016).

10 Para ver las ventajas y desventajas de los sistemas de excepciones denominados "cerrados", ver Sirinelli, P., Excepciones y Limitaciones al Derecho de Autor y los Derechos Conexos, presentado en el marco del Taller sobre Cuestiones de Aplicación del Tratado de la Ompi sobre Derecho de Autor (WCT) y el Tratado de la OMPI sobre Interpretación o Ejecución y Fonogramas (WPPT), Ginebra, 6 y 7 de diciembre de 1999.

deberán analizarse en cada caso en particular, esta excepción amplia prevista en el derecho anglosajón es la reconocida doctrina del *fair use*. Esta diferencia sustancial entre ambos sistemas cobra especial relevancia en los debates que involucran a la IA, siendo que como se mencionará más adelante, la defensa de los desarrolladores de los sistemas de IA suele enfocarse en las excepciones.

UNA PARTICULARIDAD DEL SISTEMA DE COPYRIGHT: EL *FAIR USE* Y EL CONCEPTO DE *TRANSFORMATIVE USE*

Como hemos mencionado, el sistema de Copyright anglosajón incluye una excepción clave conocida como *fair use* (uso justo), que permite el uso de obras protegidas por derechos de autor sin la necesidad de obtener permiso del titular. Esta excepción se crea para equilibrar los derechos de los creadores con el interés público, permitiendo usos como la crítica, la parodia, la educación y la investigación. Sin embargo, la aplicación de *fair use* no es automática y depende de una evaluación detallada de factores como el propósito del uso, la naturaleza de la obra protegida, la cantidad de la obra utilizada y su impacto en el mercado. En la era de la IA, esta excepción cobra especial relevancia. Varios modelos de lenguaje de gran tamaño –LLM, por sus siglas en inglés–, que se caracterizan por tener la capacidad de procesar y generar contenido utilizando grandes cantidades de información, están siendo cuestionados por el uso de obras protegidas en sus procesos de entrenamiento sin una autorización explícita, mientras sus desarrolladores alegan que dicho uso se encuentra amparado por la excepción del *fair use*.

Uno de los conceptos principales dentro del *fair use*, el cual puede resultar determinante a la hora de analizar la aplicación de esta excepción, es el concepto de *transformative use* (uso transformativo). Introducido por el caso *Campbell v. Acuff-Rose* de 1994, el *transformative use* se refiere a situaciones en las que el uso de una obra protegida aporta un nuevo propósito o significado, transformando el mensaje de la obra original de manera significativa. Este concepto puede jugar un papel crucial en la era de la IA, ya que herramientas como los mencionados LLM –en particular los modelos de IA generativa–, aunque basados en datos protegidos por derechos de autor, pueden estar generando un contenido que no replica el original, sino que lo reutiliza de manera transformadora para propósitos como el análisis, la asistencia creativa o la educación. Asimismo, se ha argumentado que el uso para fines de entrenamiento de estos sistemas es transformativo en sí mismo ya que se aleja sustancialmente del propósito original de la obra. Determinar si estos usos cumplen con el estándar de *transformative*

use será clave para resolver las controversias legales que surgen alrededor de los modelos de IA generativa.

III. PREOCUPACIONES Y PREGUNTAS LEGALES EN TORNO A LA IA Y EL DERECHO DE AUTOR

Los avances de los sistemas de IA generativa de los últimos años traen aparejados nuevas preocupaciones y preguntas legales en torno a su funcionamiento, sus capacidades y aplicaciones, particularmente respecto a las implicancias de su uso en el proceso creativo y su articulación con el derecho de autor. Los modelos de IA generativa se fundamentan en mecanismos de aprendizaje automático (*machine learning*) y emplean algoritmos de redes neuronales artificiales, lo que les permite procesar grandes volúmenes de datos, identificar patrones y, finalmente, aprender de la información con la que son entrenados.

El entrenamiento de un algoritmo con estas características necesita de grandes cantidades de datos, los cuales en muchos casos son obtenidos directamente de Internet a través prácticas como la denominada '*web scraping*', técnica que permite la recolección y extracción de información disponible en la web. Una vez entrenados, estos sistemas pueden ser utilizados para generar texto, imágenes, audio o video. La diferencia que presenta esta tecnología con la IA tradicional que conocíamos hasta hace unos pocos años es sustancial. Aquellos sistemas analizan datos y mejoran procesos a través de la automatización, para lo cual pueden realizar solamente las tareas para las que fueron diseñados siguiendo las reglas predeterminadas al momento de su programación. La IA tradicional, a diferencia de la IA generativa, no tiene capacidad de aprender. "La IA tradicional puede analizar datos y decirte lo que ve, pero la IA generativa puede usar esos mismos datos para crear algo completamente nuevo".[11]

Este es justamente uno de los aspectos que ha generado amplias discusiones últimamente. ¿La IA generativa crea o simplemente genera contenido? Si una obra no fue creada por un artista humano, ¿sigue siendo arte?[12] Hay quienes

11 Bernard Marr, "Cómo se puede explicar en forma sencilla la diferencia entre la Inteligencia artificial generativa y la tradicional", *Forbes Argentina*, 24 de julio de 2023, https://www.forbesargentina.com/innovacion/como-puede-explicar-forma-sencilla-diferencia-inteligencia-artificial-generativa-tradicional-n37661

12 Esta es la pregunta que se pregunta a profesores Harvard en el Harvard Gazette, August 15, 2023. Mineo, Liz, "If it wasn't created by a human artist, is it still art?" Consultado 30

consideran que la IA generativa tiene la capacidad de crear, mientras que hay otros que lo niegan rotundamente. La realidad es que aún no hay consenso al respecto, y esta cuestión sigue siendo un tema que invita al debate.

Un tribunal en Washington D.C. dio su respuesta a este debate, dictaminando en 2023 que una obra creada por IA sin intervención humana no puede tener derechos de autor según la ley estadounidense. El caso surgió cuando Stephen Thaler intentó registrar derechos de autor para una obra generada por su algoritmo "Creativity Machine", declarando en la solicitud al propio sistema como autor. Tras múltiples rechazos de la Oficina de Derechos de Autor, Thaler presentó una demanda, pero el tribunal federal respaldó la decisión de la Oficina.[13] Este fallo establece un precedente importante sobre los límites legales que hoy poseen las obras generadas por IA en un contexto de creciente uso de herramientas como ChatGPT, DALL-E y Midjourney.

Sin perjuicio de esto, el uso de herramientas de IA en los procesos creativos genera asimismo otro tipo de discusiones respecto a las consecuencias que pueden tener a mediano y largo plazo en la creatividad humana y el proceso de creación, habiendo quienes alegan que puede fomentar la procrastinación, producir pérdida de memoria,[14] y hasta "matar la creatividad y el pensamiento crítico".[15]

La realidad es que la IA –al menos como la conocemos hoy– no funciona de manera aislada, sino que se utiliza en el marco de un flujo de trabajo, como parte del proceso creativo, necesitando aún de la revisión y retroalimentación de una persona sobre el *output*.[16] Considerando que estos modelos aún están 'aprendiendo' y no son infalibles, los resultados obtenidos pueden ser o no

noviembre, 2024 https://news.harvard.edu/gazette/story/2023/08/is-art-generated-by-artificial-intelligence-real-art/

13 Thaler vs. Perlmutter, 2023.

14 Muhammad Abbas, Farooq Ahmed Jam, y Tariq Iqbal Khan, "Is It Harmful Or Helpful? Examining The Causes And Consequences Of Generative AI Usage Among University Students", International Journal Of Educational Technology In Higher Education 21 (16 de febrero de 2024), https://doi.org/10.1186/s41239-024-00444-7.

15 Yogesh K. Dwivedi et al., «Opinion Paper: "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives On Opportunities, Challenges And Implications Of Generative Conversational AI For Research, Practice And Policy», International Journal Of Information Management 71 (agosto de 2023): 102642, https://doi.org/10.1016/j.ijinfomgt.2023.102642. La traducción es propia.

16 Nantheera Anantrasirichai y David Bull, "Artificial Intelligence In The Creative Industries: A Review", Artificial Intelligence Review 55, n.o 1 (2 de julio de 2021): 589-656, https://doi.org/10.1007/s10462-021-10039-7.

lo que solicitamos del sistema. Para lograr resultados que tengan algún valor estético o creativo, la intervención del agente humano es indispensable, ya sea para refinar los resultados o para distinguir lo relevante de la originalidad sin mérito.[17]

Como hemos indicado las preguntas legales que surgen en relación al derecho de autor suelen dividirse en dos grupos: aquellos asuntos relacionados con el *input* (datos de entrenamiento), y aquellos que emanan de los resultados generados por los propios sistemas de IA u *output*. Las primeras se centran en el uso de obras protegidas por derecho de autor para el entrenamiento de los algoritmos, ¿es necesario contar con autorización de los titulares de derechos para usar las obras con esos fines? ¿o ese uso podría interpretarse como un uso justo aplicando la doctrina del *fair use* anglosajón? ¿Deberían ser compensados los autores por dicho uso? ¿Permiten los sistemas sustraer obras de los datos de entrenamiento cuando el autor se manifieste en contrario? Las segundas tienen que ver con los resultados que producen los algoritmos de IA generativa y sus capacidades y derivan de la ya planteada pregunta respecto a si la IA generativa puede 'crear': ¿los resultados arrojados por los sistemas de IA generativa podrían considerarse obras originales o derivadas? ¿Deberían contar estas con protección de derecho de autor? En tal caso, ¿sobre quién recaerían los derechos? ¿Quién sería el titular de esas obras?

A continuación, exploraremos varias de estas preguntas y analizaremos algunos de los casos judiciales en curso que podrían comenzar a esclarecer estas cuestiones. También abordaremos los proyectos de ley relevantes, que ofrecen indicios sobre la dirección que podría tomar la legislación en este ámbito.

IV. EL USO DE IA EN LAS INDUSTRIAS CREATIVAS. ANÁLISIS DE CASOS.

El uso de tecnologías basadas en IA en las industrias creativas no es un fenómeno reciente. Sin embargo, los modelos de IA generativa que afloraron en los últimos años han producido un gran impacto en el arte y el mundo del entretenimiento. La versatilidad de estas herramientas les ha permitido adentrarse en cada uno de los rubros creativos y, como es de esperar con una tecnología tan revolucionaria como esta, ha generado cierta conmoción entre quienes partici-

17 Este concepto ha sido traducido como "originalidad sin sentido" o "originalidad extravagante", y refiere a creaciones que si bien son nuevas u originales, no poseen valor artístico o estético.

pan de tales ámbitos. Quienes están a favor del uso de estas tecnologías suelen enfocarse en la posibilidad de democratización del arte que presentan, siendo que en particular los modelos de IA generativa permiten que cualquier persona participe de la escena creativa. En cambio, quienes se oponen alegan que estas tecnologías están ocupando el lugar de los artistas en las distintas industrias, afectando su capacidad de trabajar e incluso hasta su sustento económico, lo cual se ve agravado por la supuesta violación de sus derechos que representa el uso no autorizado de obras protegidas por derecho de autor en el proceso de entrenamiento de la IA generativa.

El temor de muchos trabajadores creativos, como guionistas, y actores de voz e intérpretes, es que las empresas utilicen sus obras o voces para entrenar modelos de IA sin ofrecer una compensación justa. En lugar de contratar por ejemplo a un intérprete de voz, o actor vocal para cada trabajo específico, las empresas podrían pedirles que graben un material mínimo a bajo costo y luego utilizar ese contenido de forma indefinida para generar nuevas producciones. Esto plantea un riesgo real de desplazamiento laboral y de explotación.[18]

En algunos rubros, en especial en la música y el ámbito editorial, estas discusiones se han trasladado al plano judicial. Por el momento, pareciera que estas disputas se han concentrado en Estado Unidos, lo cual tiene sentido, no solamente por las ya planteadas diferencias entre el Copyright anglosajón y el derecho de autor, sino porque las empresas titulares de estos programas suelen estar basadas en el territorio norteamericano. A continuación, se analizarán los principales casos judiciales que se están disputando en materia de derecho de autor e IA generativa en relación con distintos tipos de obras. La resolución de estos casos podría clarificar los límites del *fair use* y del *transformative use* en el contexto de la IA, un área que aún carece de precedentes legales específicos.

OBRAS AUDIOVISUALES

El uso de IA en la producción audiovisual –cine, televisión, distribución de contenidos en línea, publicidad e incluso videojuegos– es habitual hace varios años. Hoy en día la IA forma parte de incontables proyectos, desde el desarrollo de guión hasta su uso en postproducción, esta tecnología tiene numerosas aplicaciones.

18 Rachel Reed, "ChatNYT", Harvard Law Today, 22 de marzo de 2024, https://hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/. Consultado 29 noviembre, 2024.

Durante el año 2023 el uso de estas herramientas generó revuelo en Hollywood, tanto entre guionistas como actores, lo cual resultó en un cese en la actividad de la industria durante varios meses. Luego de una extensa huelga de ambos gremios, se llegó a un acuerdo con la Alianza de Productores de Cine y Televisión (AMPTP) y se agregaron a los contratos modelos ciertas salvaguardas para evitar el abuso de la IA en la producción y postproducción de proyectos audiovisuales. En América Latina están comenzando a surgir también casos de aplicación de estas tecnologías en películas. Por ejemplo, en el año 2023 en el Festival Internacional de Cine Documental de la Ciudad de México (DocsMX) se presentó un documental sobre la videógrafa mexicana Paola Weiss el cual mediante la implementación de herramientas de IA generativa fue narrado de manera póstuma por la propia artista. Para obtener esos resultados se trabajó con grabaciones de entrevistas y varios modelos de IA.[19] Por su parte, el cineasta mexicano Lex Ortega, reconocido en el ámbito del terror, lanzó en 2023 un corto animado realizado enteramente con un software de IA generativa que le permitía convertir texto en video.

Si bien gran parte de la industria se resiste aún a estos cambios, no observamos como sucede en otros ámbitos gran cantidad de reclamos por infracciones de derecho de autor en materia de obras audiovisuales. Por el momento solo hay un par de demandas que se centran en el uso no autorizado de material audiovisual o derechos relacionados a un producto audiovisual.

ALCON ENTERTAINMENT, LLC V. TESLA, INC

En octubre de 2024 la reconocida productora Alcon Entertainment, LLC presentó una demanda contra Tesla, Inc., Elon Musk y Warner Bros. Discovery, Inc. por infringir derechos de autor relacionados a la película 'Blade Runner 2049' de 2017.[20] La demanda tiene elementos de derecho de autor y derecho marcario, y se enfoca en el uso no autorizado de capturas o *stills* de la película para solicitar a un algoritmo de IA la creación de una nueva imagen que sea similar a aquellas, imagen que es identificada por el demandante como una obra derivada no autorizada de la película.[21]

19 César Huerta Ortiz, "Inteligencia Artificial Pone a Prueba a Filmes Mexicanos," *El Universal*, October 13, 2023, https://www.eluniversal.com.mx/espectaculos/ia-pone-a-prueba-a-filmes-mexicanos/.

20 Alcon Entertainment, LLC v. Tesla, Inc., 2:24-cv-09033, (C.D. Cal.), https://storage.courtlistener.com/recap/gov.uscourts.cacd.945374/gov.uscourts.cacd.945374.1.0.pdf

21 El actor manifiesta que los demandados habían solicitado una licencia para utilizar una

MILLETTE V. OPENAI[22]; MILLETTE V. GOOGLE[23] Y MILLETTE V. NVIDIA[24]

El youtuber David Millette presentó en agosto de 2024 tres demandas contra distintos titulares de algunos de los principales sistemas de IA generativa: OpenAI titular de ChatGPT, Google titular de Gemini y NVIDIA titular de Cosmos AI. Las demandas –presentadas como acciones de clase en representación de todos aquellos que se encuentren en una situación similar– alegan que los acusados han estado transcribiendo de forma encubierta vídeos de YouTube obtenidos mediante la técnica de *web scraping* para crear conjuntos de datos de entrenamiento que luego utilizan para entrenar sus productos de IA.[25] Millette alega que el uso que están realizando los sistemas de IA de sus videos no ha sido autorizado por los titulares de derecho, es decir, los propios creadores, siendo que según establecen los Términos de Uso de la plataforma YouTube los usuarios conservan los derechos de propiedad intelectual sobre el material que suben.

Inicialmente las tres demandas contemplaban reclamos fundados en cuestiones de enriquecimiento ilícito y competencia desleal, sin embargo, en el mes de octubre de 2024 se presentó una adenda a la demanda inicialmente presentada contra OpenAI en donde no solamente se sumó una co-demandante, Ruslana Petryazhna, sino que además se incorporó al reclamo una infracción a los derechos de autor de la Sra. Petryazhna. En el escrito incorporado se hace referencia a que las obras de la Sra. Petryazhna disponibles en YouTube están debidamente registradas ante la Oficina de Copyright de los Estados Unidos, lo cual pareciera indicar que la ausencia inicial de menciones a violaciones de

captura de la película en un evento publicitario de la empresa Tesla para el lanzamiento del auto autónomo 'Cybercab' –evento que se transmitió en vivo por la plataforma X– y la productora no solo se había negado rotundamente a su uso, sino que además se había opuesto a cualquier asociación directa o indirecta de su obra con la empresa de Elon Musk, su persona o el producto en cuestión. La actora afirma que las imágenes de la secuencia que fueron tomadas de referencia para la generación de la nueva imagen forman parte del "núcleo creativo" o "corazón" de la obra y por lo tanto son elementos protegibles mediante Copyright.

22 Millette v. OpenAI, Inc. (5:24-cv-04710), https://www.courtlistener.com/docket/69012680/millette-v-openai-inc/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

23 Millette v. Google LLC (5:24-cv-04708) https://www.courtlistener.com/docket/69012654/millette-v-google-llc/?order_by=desc

24 Millette v. Nvidia Corporation, 5:24-cv-05157 https://www.classaction.org/media/millette-v-nvidia-corporation.pdf

25 Millette v. OpenAI; Millette v. Google.

derechos de autor en los reclamos ingresados por Millette se debía a otra de las diferencias cruciales entre el Copyright y el derecho de autor: la exigencia de formalidades para la existencia de la protección.

Respecto a la licitud de la implementación de la técnica de *web scraping* usada sobre el contenido disponible en la plataforma de videos, el director ejecutivo de YouTube, Neal Mohan, en una entrevista con Bloomberg manifestó expresamente que los Términos de Servicio de la plataforma permiten que se obtenga cierta información mediante *scraping*, como el título de los videos, el nombre de los canales o del propio creador, porque así es como se habilita a que esos datos luego aparezca por ejemplo en motores de búsqueda, pero expresamente aclaró que no permiten que se descarguen transcripciones o fragmentos de videos, lo cual constituiría una clara violación de sus políticas de uso.[26]

Esta es una de las demandas que OpenAI ha solicitado se transfiera a un foro de Litigio Multidistrital (MDL, por sus siglas en inglés). La transferencia de las demandas relacionadas con derechos de autor en IA de los tribunales del Distrito Norte de California y el Distrito Sur de Nueva York, a un foro único permitiría centralizar y coordinar la recopilación de pruebas y todos los procedimientos previos al juicio, incluidos los fallos sumarios. En cuanto al fallo sumario, el MDL evitaría posibles resultados inconsistentes sobre la cuestión del *fair use* en el entrenamiento de modelos de IA. Si el tribunal del MDL se niega a emitir juicio sumario sobre este tema, la cuestión del uso justo tendría que ser juzgada en los tribunales originales durante el juicio.[27]

OBRAS MUSICALES

La IA también plantea desafíos para los derechos de autor y los derechos conexos en la industria musical. Como se ha adelantado, el uso de obras preexistentes como datos de entrenamiento puede infringir derechos de autor si se realiza sin autorización. Las herramientas de IA tienen la capacidad de generar canciones completas o fragmentos imitando composiciones o estilos específicos. El clonado de voces o los *deepfakes* plantea preocupaciones sobre

26 Bloomberg Television, "YouTube Says Using Videos to Train OpenAI's Sora Breaks Rules", 4 de abril de 2024. Video, 0:00:36, https://www.youtube.com/watch?v=FBZ__BeChRg

27 Annelise Levy, Aruni Soni, y Kyle Jahner, "OpenAI Bid to Centralize IP Suits Follows Discovery Setbacks", *Bloomberg Law*, 6 de diciembre de 2024, https://news.bloomberglaw.com/ip-law/openai-bid-to-centralize-ip-suits-follows-discovery-setbacks. Consultado 6 diciembre, 2024.

el uso no autorizado de las voces de los artistas y el potencial impacto en sus derechos de imagen y propiedad intelectual. Adicionalmente, la creación de contenidos derivados o de calidad similar a obras protegidas podría diluir el valor de los derechos conexos, como los de interpretación o grabación.

La tendencia a la imitación no autorizada de artistas y canciones[28] ha sido impulsada en gran parte por los sistemas musicales de IA disponibles que ofrecen modelos *freemium*[29] cuyo uso resulta sencillo e intuitivo, por lo que cualquier persona puede utilizarlos. Aunque esto ofrece la oportunidad de componer música a más personas, aún sin tener conocimientos previos, su uso indebido puede tener consecuencias perjudiciales. Las voces y estilos de varios artistas de renombre internacional han sido recreados mediante IA en composiciones musicales y grabaciones que imitan de manera notable las características distintivas de sus interpretaciones, sin que los artistas hayan participado en su creación ni otorgado su consentimiento para su uso. En abril de 2023 se viralizó una canción llamada "Heart on My Sleeve" que replicaba las voces de los artistas Drake y The Weekend, la cual había sido enteramente generada utilizando IA. La canción logró obtener millones de reproducciones entre TikTok, YouTube y Spotify[30] antes de que la empresa Universal Music Group –de la que forman parte ambos artistas– lograra que se bajara de las plataformas. Otro artista afectado por estos abusos es el cantante puertorriqueño conocido como Bad Bunny, quien no ha tomado a la ligera el uso de su voz sin su consentimiento manifestándose efusivamente en contra de una canción creada con IA que circulaba en la plataforma TikTok.[31]

Gran parte de los programas de IA que permiten generar estos materiales están basados en modelos de aprendizaje automático o *machine learning* y su uso para la creación de nuevas obras musicales profundiza las preguntas respecto al uso no autorizado de obras protegidas en el entrenamiento de estos algoritmos. En Estados Unidos actualmente hay varias disputas iniciadas tanto

28 Isabela Raygoza, "Bad Bunny Critica Canciones Que Lo Imitan Con IA Que Se Hacen Virales En TikTok", *Billboard*, 11 de septiembre de 2023, https://www.billboard.com/espanol/musica/bad-bunny-critica-canciones-generadas-ia-inteligencia-artificial-1235467073/.

29 El modelo freemium obtiene su nombre de la conjunción de los términos en inglés 'free' y 'premium' y refiere a un modelo de negocios que consiste en ofrecer a los usuarios un servicio básico de manera gratuita y cobrar por algunas funciones u opciones avanzadas o premium.

30 Joe Coscarelli, "An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World", *The New York Times*, 19 de abril de 2023, https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html.

31 Isabela Raygoza, 2023.

por sellos discográficos como por editoras musicales contra empresas titulares de software con estas características.

UMG RECORDINGS, INC. V. SUNO, INC. Y UMG RECORDINGS V. UNCHARTED LABS

Universal Music Group, Warner Music Group y Sony Music Group, entre otros, presentaron dos demandas alegando infracción de Copyright de obras en sus repertorios. En el escrito inicial[32] de la demanda presentada contra la empresa Suno, Inc., titular del programa homónimo, las discográficas alegan que para obtener resultados que se asemejen tanto a las obras originales, los modelos de IA deben haber utilizado las obras protegidas para su entrenamiento, uso que no ha sido autorizado por los titulares de derecho. Incluso los actores presentan varios ejemplos en los cuales al proveer ciertos *prompts* o indicaciones al sistema, este generó resultados que contienen grandes similitudes melódicas a las obras de su titularidad, dejando en evidencia que la empresa utilizó esas obras en su proceso de entrenamiento.

En el escrito de contestación, la empresa Suno explica cómo funciona el sistema[33] y alega que se tomaron ciertos recaudos para evitar que los resultados arrojados se parezcan a las obras a las que se expuso el modelo en su proceso de entrenamiento –como por ejemplo, no permitir el proceso de solicitudes que incluyan el nombre de un artista en particular–.

El núcleo del debate, como en muchas otras demandas, gira en torno a si el uso de obras protegidas por derechos de autor para entrenar modelos de IA puede considerarse *fair use* o uso justo. Los actores, anticipándose a la postura del demandado, rechazan que dicho uso pueda encuadrarse dentro de esta excepción. Por su parte, el demandado argumenta que la copia de obras protegidas con fines de entrenamiento no constituye una infracción, amparándose en precedentes legales de tribunales estadounidenses que han permitido realizar copias 'inmediatas' e invisibles para los usuarios, siempre que estas sean utilizadas para generar resultados no infractores. Suno incluso refuta los planteos de los actores respecto al cuarto factor del *fair use* –efecto en el mercado potencial de la obra– citando casos como los de Google LLC v. Oracle Am., Inc., (2021) y Sega Enters. Ltd. v. Accolade, Inc. (1992).

32 UMG Recordings, Inc. v. Suno, Inc., 1:24-cv-11611, (D. Mass.), https://storage.courtlistener.com/recap/gov.uscourts.mad.272063/gov.uscourts.mad.272063.1.0.pdf

33 Suno es un programa que permite a las personas utilizar descripciones de géneros, estilos y otros elementos musicales en inglés sencillo (no técnico) para crear obras musicales.

En paralelo se interpuso asimismo una demanda similar contra Uncharted Labs, Inc.,[34] la empresa responsable de Udio, otro software de IA generativa con características comparables a Suno. Tanto los fundamentos y argumentos de la demanda como la respuesta presentada por la empresa son prácticamente idénticos a los del caso previamente mencionado.

CONCORD MUSIC GROUP, INC. V. ANTHROPIC PBC

Las editoriales musicales más importantes del mercado también iniciaron un reclamo por el uso no autorizado de las composiciones musicales de sus representados en el entrenamiento de un modelo de IA generativa denominado "Claude", de la empresa Anthropic PBC. A diferencia de los sistemas antes mencionados, Claude es un '*chatbot*', un LLM de propósito general que permite generación de texto y código. Los actores alegan que Anthropic recurre a la práctica de *web scraping* para entrenar su algoritmo. La infracción principal está en el uso de estas obras como parte del *input*, ya que Anthropic habría copiado las obras para procesarlas violando así los derechos de autor de sus titulares.

Adicionalmente, los actores manifiestan haber detectado que los usuarios pueden solicitar y obtener del sistema *Claude* copias textuales o casi textuales de las letras de las obras de su titularidad, alegando que los resultados arrojados por este son sustancial y sorprendentemente similares a las obras originales protegidas por derechos de autor.[35] En concreto, en lo que respecta al *output,* los actores detectan dos tipos de infracciones; por un lado, cuando se le consulta al sistema por la letra de una canción este arroja respuestas que replican casi literalmente las letras originales, lo cual deja en evidencia que el modelo ha tenido acceso a esas obras y las utiliza sin autorización cuando en realidad el demandado debería haber obtenido una licencia para tales fines. Por otro lado, los actores alegan que *Claude* "es capaz de generar letras para nuevas canciones que incorporan las letras de canciones protegidas por derechos de autor. En estos casos, la producción de Claude puede ser que incluya partes de una obra protegida junto con partes de otras obras protegidas de una manera que es completamente inconsistente e incluso contraria a la forma en

34 UMG Recordings, Inc. v. Uncharted Labs, Inc. (1:24-cv-04777), https://www.courtlistener.com/docket/68878697/umg-recordings-inc-v-uncharted-labs-inc/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

35 Concord Music Group, Inc. v. Anthropic PBC, 5:24-cv-03811, (N.D. Cal.), https://storage.courtlistener.com/recap/gov.uscourts.tnmd.96652/gov.uscourts.tnmd.96652.1.0.pdf. La traducción es propia.

que el compositor las concibió",[36] incluso cuando esto no sea específicamente solicitado por el usuario en los *prompts* introducidos.

Es interesante la observación de los actores que, en ciertos casos, el sistema se niega a proveer determinada información debido a normas de Copyright. Por ejemplo, cuando se le solicita las letras de determinadas obras musicales, el propio programa responde que hacerlo podría afectar los derechos de autor de terceros. Esto deja en evidencia que la demandada posee la capacidad de implementar ciertas 'barreras' para evitar –al menos parcialmente– las infracciones detectadas por el grupo de editoriales.

Luego de transcurrido un año desde su inicio, los actores solicitaron una medida cautelar para que el demandado no continúe infringiendo los derechos de autor de las editoriales representadas en la demanda hasta tanto la cuestión se resuelva. Esta es la primera demanda contra una empresa de IA en la cual se presenta una medida cautelar de este estilo. En la medida presentada los actores solicitan que Anthropic: (i) aplique y mantenga medidas efectivas para evitar que sus modelos de IA generen resultados que reproduzcan, distribuyan o muestren las letras de las obras de su titularidad; (ii) aplique y mantenga medidas efectivas para evitar la creación de obras derivadas de obras de su titularidad; (iii) se abstenga de utilizar obras protegidas de titularidad de los actores para entrenar futuros sistemas de IA. Por supuesto, la demandada rechazó esta medida, explicando que el uso de obras protegidas en el entrenamiento de sus modelos recae dentro de los parámetros de la doctrina del *fair use* y por lo tanto no infringe los derechos de los actores. La demandada hace hincapié en que *Claude* no es un *chatbot* de letras de canciones en particular, sino que por la forma en que funcionan los sistemas como el que se juzga, es decir, LLMs de propósito general, puede ser que se hayan incluido obras de titularidad de los reclamantes entre los datos de entrenamiento mas "[e]l propósito de incluir instancias de [esas obras] en el conjunto de datos de entrenamiento no es, en última instancia, replicar la forma particular en que cualquier autor original expresó la idea incorporada en el texto, que es lo que protege la ley de derechos de autor",[37] por lo tanto "[c]ualquier inclusión de las letras de las canciones de los demandantes… sería simplemente un subproducto del único enfoque viable para resolver [el] desafío técnico"[38] que presenta el entrenamiento de estos modelos.

36 Concord Music Group, Inc. v. Anthropic PBC. La traducción es propia.
37 Concord Music Group, Inc. v. Anthropic PBC. La traducción es propia.
38 Concord Music Group, Inc. v. Anthropic PBC. La traducción es propia

En su más reciente presentación en este juicio, Anthropic argumenta que el tribunal debería centrar su análisis en determinar "si el uso de materiales protegidos por derechos de autor para extraer contexto estadístico y factual con el propósito de entrenar modelos de IA generativa como *Claude* califican como un *transformative fair use* (uso justo transformativo) conforme a la legislación de Copyright".[39]

Este planteo de Anthropic expone el núcleo del debate: ¿el uso de material protegido por derechos de autor para entrenar sistemas de IA generativa, que posteriormente producen obras nuevas, puede considerarse un *transformative use* según la jurisprudencia estadounidense? Este es el *quid* de la cuestión que los tribunales deberán resolver, ya que el concepto de 'uso transformativo' exige que el nuevo uso aporte un propósito, carácter, mensaje o valor distintivamente nuevo y creativo en relación con el material original.

OBRAS LITERARIAS

Uno de los ámbitos más afectados por el auge de la IA generativa ha sido el literario. Desde novelas hasta noticias periodísticas, los sistemas de IA se han apropiado de parte del mercado literario. Durante el año 2023 las tiendas virtuales de Amazon se vieron inundadas de nuevas publicaciones escritas que acreditaban a ChatGPT como autor, tanto así que la empresa se vio forzada a imponer un límite a la cantidad diaria de obras generadas con IA que se podrían crear por día en su servicio de '*self-publishing*' Amazon Kindle Direct Publishing (KDP).[40]

Mary Rasenberger, directora ejecutiva del Sindicato de Autores de Estados Unidos (*Authors Guild*), ha advertido sobre el impacto que puede tener esta tendencia en el trabajo de los autores, haciendo hincapié en la importancia de transparentar cuando las obras han sido creadas por IA.[41] Actualmente, los

39 Chris Cooke, «Anthropic's Latest Legal Filing Tells Court To Focus On Whether Using Lyrics Is "Transformative Fair Use"», Complete Music Update, 19 de agosto de 2024, https://completemusicupdate.com/anthropics-latest-legal-filing-tells-court-to-focus-on-whether-using-lyrics-is-transformative-fair-use/. La traducción es propia. Consultado 10 diciembre, 2024.

40 Kindle Direct Publishing Community. "Update on KDP Title Creation Limits", 18 de septiembre de 2023.https://www.kdpcommunity.com/s/article/Update-on-KDP-Title-Creation-Limits?language=en_US&forum=KDP%20Forum.

41 Greg Bensinger, "Focus: ChatGPT launches boom in AI-written e-books on Amazon", *Reuters*, 21 de febrero de 2023, https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/.

términos de uso de KDP[42] establecen la obligación de informar sobre el contenido generado por IA incluido en la publicación de un nuevo libro o una nueva edición de una obra ya existente, ya sea que se trate de texto, imágenes o traducciones, incluyendo incluso los casos en los cuales la persona realizó ediciones sustanciales a dicho contenido. En cambio, no es necesario informar sobre el contenido 'asistido por IA', es decir, si se emplearon herramientas basadas en IA para generar ideas, realizar ediciones, corregir errores o mejorar la calidad de la obra final.

Se identifican varios problemas en el uso de IA generativa en la escritura. Principalmente el abuso de estas herramientas puede ser perjudicial para la creatividad y originalidad, siendo que los contenidos generados por estos sistemas pueden eventualmente resultar repetitivos. Estos programas muchas veces quedan en evidencia dado al uso excesivo de ciertas palabras o por incluir recursos estilísticos que llaman la atención del lector y elevan sospechas de que no se trata de una escritura humana genuina.[43]

En el ámbito literario y editorial, han surgido controversias importantes relacionadas con el uso no autorizado de obras protegidas por derechos de autor por parte de sistemas de IA. Varias editoriales y titulares de derechos han presentado reclamos legales, argumentando que estas tecnologías utilizan sus contenidos sin consentimiento ni compensación adecuada. Asimismo, han surgido un gran número de demandas por parte de medios de comunicación presentados contra titulares de programas de IA generativa, destacándose entre ellos el iniciado por el periódico The New York Times contra OpenAI y Microsoft. Sin perjuicio de lo anterior, se debe destacar que una gran cantidad de empresas del ámbito periodístico y editorial –entre ellas Reuters, Associated Press y Time–[44] han decidido tomar otro camino y firmar acuerdos con compañías de IA permitiéndoles el acceso a su historial de publicaciones.

---

42 "Content Guidelines", Kindle Direct Publishing, s. f., https://kdp.amazon.com/en_US/help/topic/G200672390#:~:text=We%20require%20you%20to%20inform,to%20disclose%20AI%2Dassisted%20content.

43 Tony Berber Sardinha, "AI-generated Vs Human-authored Texts: A Multidimensional Comparison", *Applied Corpus Linguistics* 4, n.o 1 (20 de diciembre de 2023): 100083, https://doi.org/10.1016/j.acorp.2023.100083.

44 Charlotte Tobitt, "Who's suing AI and who's signing: Publisher deals vs lawsuits with generative AI companies", *Press Gazette*, 6 de diciembre de 2024, https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/.

THE NEW YORK TIMES V. MICROSOFT Y OPENAI[45]

The New York Times ("NYT") demandó a OpenAI y Microsoft por el uso no permitido de artículos del NYT para entrenar sus sistemas de IA. El caso podría tener un impacto significativo en la relación entre la IA generativa y los derechos de autor, particularmente con respecto al mencionado *fair use*, y podría, en última instancia, determinar cómo se construyen, entrenan y usan los modelos de IA generativa y otros LLMs.

La demanda del NYT se centra en el uso de obras protegidas por derechos de autor en el desarrollo de herramientas de IA generativa, como ChatGPT de OpenAI y Bing Chat de Microsoft, ambos basados en modelos GPT[46] –Generative Pre-trained Transformer o Transformador Generativo Pre-entrenado en español–. Estas herramientas se entrenan con vastos conjuntos de datos textuales para aprender patrones de palabras en contextos específicos. Los modelos más recientes utilizan miles de millones de palabras, equivalentes a "un documento de Microsoft Word de más de 3.700 millones de páginas".[47] El NYT alega que estos datos incluyen una gran cantidad de contenido protegido por sus derechos de autor.[48] La demanda también subraya el daño potencial a la marca del Times a través de las llamadas 'alucinaciones' de la IA, un fenómeno por el cual los *chatbots* dan información falsa a los usuarios.

El NYT acusa a OpenAI de infringir derechos de autor al usar y reproducir sus obras sin licencia durante el entrenamiento de sus modelos. El problema se agrava porque, primero, los LLM a veces 'memorizan' partes del contenido de entrenamiento, generando reproducciones casi textuales. Segundo, producen resultados de búsqueda 'sintéticos' que pueden ofrecer contenido más detallado de un artículo original que una búsqueda tradicional, permitiendo a los usuarios eludir las suscripciones pagas que el NYT pretende cobrar a sus lectores.

La herramienta "Browse With Bing" de ChatGPT ha generado controversia debido a su forma de presentar los resultados de búsqueda, especialmente en el caso de Wirecutter, el reconocido sitio de recomendaciones de productos

45 The New York Times Company v. Microsoft Corporation (1:23-cv-11195), https://www.courtlistener.com/docket/68117049/the-new-york-times-company-v-microsoft-corporation/

46 Se trata de modelos basados en aprendizaje automático que tienen la capacidad de comprender y generar texto en lenguaje natural.

47 Audrey Pope, "NYT V. OpenAI: The Times's About-Face", Harvard Law Review, 10 de abril de 2024, https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/. Consultado 28 de noviembre 2024. La traducción es propia.

48 Pope (2024). La traducción es propia.

del NYT. Esta herramienta reproduce casi literalmente el contenido de Wirecutter, pero elimina los enlaces originales que el sitio utiliza para dirigir a los usuarios hacia los productos recomendados. Estos enlaces son fundamentales para el modelo de negocio de Wirecutter, ya que generan comisiones por las ventas derivadas de sus recomendaciones. Al omitir estos enlaces, "Browse With Bing" no solo impide que Wirecutter reciba ingresos, sino que también plantea serias preocupaciones sobre el uso justo y ético del contenido protegido.

Según el NYT este caso supone una amenaza grave para el periodismo de calidad. Alegan que millones de artículos se usaron para entrenar a los *chatbots* de las demandadas. ChatGPT ofrece a sus usuarios fragmentos casi textuales de artículos de NYT que la demandante sólo ofrece con una suscripción paga. Si los lectores pueden acceder a reproducciones casi textuales de sus artículos de manera gratuita mediante estos programas, se reducirá la necesidad de suscribirse al diario, afectando su capacidad de financiar el periodismo. Esto, advierten, tendría un alto costo para la sociedad, ya que "el periodismo independiente es vital para nuestra democracia y es cada vez más raro y valioso".[49] En contraste, describen a OpenAI como "un negocio multimillonario con fines de lucro basado, en gran parte, en la explotación sin licencia de obras protegidas por derechos de autor".[50]

Por otro lado, OpenAI no parece disputar que los artículos del NYT formen parte de los datos utilizados para entrenar sus modelos. De hecho, ha sido relativamente transparente sobre el material empleado en el desarrollo de sus grandes modelos de lenguaje. Sin embargo, el concepto de 'similitud sustancial', básico en el análisis de los tribunales estadounidenses al analizar las infracciones al Copyright, es crucial en este caso para evaluar si se han copiado los elementos expresivos de una obra protegida por derechos de autor o únicamente las ideas contenidas en ella. Esto se debe a que los derechos de autor protegen la forma en que se expresan las ideas, pero no las ideas en sí mismas. Es perfectamente legal reinterpretar las ideas de alguien en palabras propias, sin infringir. No obstante, si se reproduce literalmente el lenguaje o la expresión de otra persona, eso sí constituye una violación.

La defensa de OpenAI, al igual que ha sucedido con otros de los casos anteriormente mencionados, se fundamenta en el principio de *fair use*. OpenAI se

49 Pope (2024). La traducción es propia.

50 No. 1:23-cv-11195 (S.D.N.Y.), citado en Baker Hotstetler Case Tracker: Artificial Intelligence, Copyrights and Class Actions, https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/

aferra al concepto de *transformative use*[51] para argumentar que su modelo GPT no reproduce el *input* de forma literal y que este tampoco es su propósito original, sino que lo transforma al analizar patrones, generar respuestas contextuales y proporcionar nuevos usos creativos y funcionales. Según OpenAI, su modelo no sustituye al contenido original, como los artículos del NYT, sino que crea una herramienta que sirve a propósitos diferentes, como la asistencia en tareas, el aprendizaje o la investigación.

Fundamentalmente, Open AI no ha refutado el uso de obras del NYT sin consentimiento. En lugar de argumentar en contra de las acusaciones de infracción *prima facie* de derechos de autor presentadas por la actora, OpenAI simplemente argumenta que tales infracciones están justificadas según la doctrina del *fair use*.[52]

DAILY NEWS LP V. MICROSOFT CORPORATION Y OPENAI[53]

En un caso similar, el periódico *Daily News* demandó a Microsoft y OpenAI en el Distrito Sur de Nueva York por infracción directa, indirecta y contributiva de derechos de autor, violaciones a la Digital Millennium Copyright Act (DMCA),[54] competencia desleal, dilución de marcas registradas y dilución y daño a la reputación empresarial. Microsoft y OpenAI han presentado mociones para desesti-

51 Como fue mencionado, el *transformative use* se estableció en 1994 en el caso Campbell v. Acuff-Rose Music, Inc. En ese caso, la Corte Suprema de Estados Unidos determinó que un uso no autorizado de una obra protegida podía considerarse *fair use* si aportaba un nuevo significado, propósito o carácter a la obra original, transformándola de manera significativa. El ejemplo específico fue una parodia de la canción "Oh, Pretty Woman" por el grupo 2 Live Crew, que la corte consideró transformadora porque reinterpretaba la canción original en un contexto humorístico y crítico.

52 Sundara Rajanm, Mira T. Is Generative AI Fair Use of Copyright Works? NYT v. OpenAI,Kluwer Copyright Blog, Consultado 28 Noviembre, 2024, https://copyrightblog.kluweriplaw.com/2024/02/29/is-generative-ai-fair-use-of-copyright-works-nyt-v-openai/#:~:text=Crucially%2C%20the%20use%20of%20NYT,the%20doctrine%20of%20fair%20use.

53 Daily News LP v. Microsoft Corporation (1:24-cv-03285), https://www.courtlistener.com/docket/68484432/daily-news-lp-v-microsoft-corporation/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

54 La Digital Millennium Copyright Act (DMCA) es una ley federal que regula los derechos de autor de Estados Unidos que implementa dos tratados del año 1996 de la Organización Mundial de la Propiedad Intelectual (OMPI).

mar algunos de los reclamos, y OpenAI ha presentado una moción de consolidación ante el New York Times.[55]

THE INTERCEPT MEDIA V. OPENAI[56] Y RAW STORY MEDIA V. OPENAI[57]

En los casos *The Intercept Media* y *Raw Story Media v. OpenAI* iniciados en febrero de 2024, un grupo de organizaciones periodísticas presentó demandas casi idénticas alegando violaciones a la mencionada DMCA. Estas denuncias se basan en la supuesta inclusión de trabajos periodísticos de los demandantes en los conjuntos de datos utilizados para entrenar ChatGPT. Los demandantes alegan que en el proceso de entrenamiento los desarrolladores deliberadamente eliminaron la información de gestión de derechos de autor (CMI) protegida por la DMCA, programando los sistemas para que no notificaran a los usuarios cuando las respuestas arrojadas estaban protegidas por los derechos de autor de los periodistas. Por su parte, los demandados han solicitado la desestimación de los reclamos, argumentando que los demandantes no han demostrado que se hayan distribuido públicamente copias de sus obras desprovistas de la CMI.[58]

55 No. 1:24-cv-03285 (S.D.N.Y), citado en Baker Hotstetler Case Tracker: Artificial Intelligence, Copyrights and Class Actions, https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/

56 The Intercept Media, Inc. v. OpenAI, Inc. (1:24-cv-01515), https://www.courtlistener.com/docket/68290804/the-intercept-media-inc-v-openai-inc/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

57 Raw Story Media, Inc. v. OpenAI Inc. (1:24-cv-01514), https://www.courtlistener.com/docket/68290709/raw-story-media-inc-v-openai-inc/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

58 Nos. 1:24-cv-01514, 1:24-cv-01515 (S.D.N.Y.), citado en Baker Hotstetler Case Tracker: Artificial Intelligence, Copyrights and Class Actions, https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/

NAZEMIAN V. NVIDIA CORPORATION[59] Y DUBUS V. NVIDIA CORPORATION[60]

En los casos *Nazemian y Dubus v. NVIDIA Corporation*, dos grupos de autores presentaron demandas colectivas relacionadas contra NVIDIA Corporation, alegando que la empresa utilizó, sin autorización, libros protegidos por derechos de autor para entrenar su modelo de lenguaje Nemo Megatron-GPT. Según los demandantes, "cada vez que un modelo de lenguaje genera texto en respuesta a una solicitud del usuario, realiza un cálculo con el objetivo de imitar la expresión protegida que fue incorporada a partir del conjunto de datos de entrenamiento". Estas demandas, que se encuentran en curso en el Tribunal del Distrito Norte de California, plantean importantes interrogantes sobre el uso de obras protegidas en el entrenamiento de programas de IA.[61]

THOMSON REUTERS V. ROSS INTELLIGENCE[62]

En mayo de 2020 Thomson Reuters, empresa matriz de Westlaw, demandó a la empresa de investigación legal basada en IA, ROSS Intelligence, alegando que esta copió de manera no autorizada contenido de la plataforma Westlaw para entrenar su propio motor de búsqueda legal basado en IA. ROSS respondió las acusaciones de la actora manifestando que los elementos que Thomson Reuters alegaba se habían copiado no son protegibles por Copyright y reconvino la demanda incorporando cuestiones relacionadas a conductas monopólicas y anticompetitivas por parte de Thomson Reuters. En diciembre de 2024 se llevó a cabo una audiencia respecto de las mociones de juicio sumario presentadas por ambas partes. Previamente el juez Stephanos Bibas recomendó a los abogados enfocarse en dicha audiencia, entre otras, en las siguientes cuestiones: (i) el

59 Nazemian v. NVIDIA Corporation (4:24-cv-01454), https://www.courtlistener.com/docket/68325563/nazemian-v-nvidia-corporation/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

60 Dubus v. NVIDIA Corporation (4:24-cv-02655), https://www.courtlistener.com/docket/68497292/dubus-v-nvidia-corporation/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

61 Nos. 3:24-cv-01454, 3:24-cv-02655 (N.D. Cal.), citado en Baker Hotstetler Case Tracker: Artificial Intelligence, Copyrights and Class Actions, https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/

62 Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc. (1:20-cv-00613), https://www.courtlistener.com/docket/17131648/thomson-reuters-enterprise-centre-gmbh-v-ross-intelligence-inc/?filed_after=&filed_before=&entry_gte=&entry_lte=&order_by=desc

primer factor del *fair use* –el propósito y el carácter del uso–; (ii) el cuarto factor del *fair use* –el efecto del uso sobre el mercado potencial–; (iii) el tipo de Copyright que Thomson Reuters posee sobre las partes de Westlaw en disputa –notas explicativas de los casos, sistema de clasificación 'Key Number System'–; y (iv) las circunstancias en las que es apropiado que un juez resuelva la similitud sustancial mediante un juicio sumario en lugar de dejarlo en manos de un jurado. Este caso iniciado en mayo de 2020 es uno de los más avanzados contra una empresa titular de IA y, si bien puede ser que la cuestión se defina en base a si el material copiado es protegible o no por Copyright, las resoluciones que se emitan pueden ser relevantes para definir cuestiones relativas a la interpretación del *fair use*, el *trransformative use* y los límites del uso de contenidos protegidos en el entrenamiento de herramientas de IA.[63]

Estos son solo algunos de los conflictos que se están disputando contra empresas titulares de IA, mas son suficiente para subrayar la creciente tensión entre la innovación tecnológica y la protección de los derechos de autor, planteando interrogantes sobre cómo equilibrar el acceso a datos con la compensación justa a los creadores y titulares de derechos.

Las causas de acción que se repiten en muchos de los casos, además por supuesto de las infracciones en materia de Copyright, incluyen violaciones de la DMCA, reclamos por incumplimiento de contrato, conductas anticompetitivas, así como una variedad de agravios. En lo que respecta a los derechos de las obras utilizadas para el entrenamiento de los modelos de IA, la defensa por defecto a la que recurren los demandados es el *fair use*, alegando que dicho uso recae dentro de esta excepción. Uno de los argumentos a los cuales se aferran los demandados tiene que ver con el *transformative use*, el cual pesa a favor de la interpretación de *fair use*.

La definición de estos casos será determinante para el avance de esta tecnología y si bien nuestro interés se centra en las disputas relacionadas a los derechos de autor, es innegable que existen cuestiones económicas de fondo que pueden afectar los resultados de estas resoluciones.

63 No. 1:20-cv-00613 (D. Del.), citado en Baker Hotstetler Case Tracker: Artificial Intelligence, Copyrights and Class Actions, https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/

V. DESAFÍOS Y LIMITACIONES ACTUALES DEL DERECHO DE AUTOR ANTE EL AVANCE DE LA IA GENERATIVA

Los casos comentados dejan en evidencia que hay una imperante necesidad de respuestas que esclarezcan la situación de las obras en la era de la IA. Los cambios introducidos por la IA generativa han sido sustanciales, impactando todo el proceso creativo. Si bien no es extraño que el derecho quede desactualizado ante grandes cambios tecnológicos, los legisladores de ninguna manera podrían haber contemplado las necesidades actuales de los autores y estas nuevas herramientas. Las normas que se aplican hoy a la relación entre obras e IA no proveen soluciones a las disputas que surgen a raíz del uso de esta tecnología. Continuar aplicando la normativa actual sobre derechos de autor en Latinoamérica en materia de IA generativa tiene ciertas implicancias como prohibir el uso no autorizado de obras en los *datasets* de entrenamiento y negar la autoría –y por lo tanto, protección– a creaciones generadas por IA. Aunque la falta de una excepción de *fair use* en las legislaciones civilistas ofrece mayor certeza jurídica frente a Estados Unidos, la rigidez del sistema de excepciones taxativas deja en evidencia las limitaciones de estos sistemas al intentar responder a las demandas de una era tecnológica con un marco legal desfasado

Ante esta situación surgen entonces dos cuestiones principales que requieren nuestra atención: por un lado con respecto al *input* se cuestiona la legalidad del uso de obras protegidas como parte del entrenamiento de los algoritmos de IA, y por el otro, con respecto al *output* se disputa la autoría del material generado por los sistemas de IA generativa.

DETERMINACIÓN DE LA LEGALIDAD DEL USO DE OBRAS PROTEGIDAS COMO *INPUT* EN EL ENTRENAMIENTO DE IA SEGÚN EL MARCO LEGAL VIGENTE

En lo que respecta al *input* lo primero que hay que determinar es si el uso de obras para el entrenamiento de sistemas de IA constituye o no un ilícito. El sistema de derechos de autor continental que rige en nuestra región otorga al autor la facultad de decidir sobre el destino de su obra, es decir que todo uso –salvo que recaiga dentro de algunas de las excepciones previstas por la ley– requiere de la previa autorización del titular de derechos.

El uso de obras protegidas para entrenar sistemas de IA no se encuentra actualmente previsto dentro de las excepciones que plantean la mayoría de

las legislaciones latinoamericanas en materia de derecho de autor, como así tampoco la recolección de obras mediante técnicas de *web scraping*. En América Latina, los países generalmente carecen de excepciones al derecho de autor que aborden metodologías de investigación, como el Text and Data Mining (TDM) y el Machine Learning (ML), técnicas esenciales para analizar grandes volúmenes de datos y entrenar modelos de IA. Según un informe de la Alianza de La Sociedad Civil Latinoamericana para el Acceso Justo al Conocimiento, de los países con estrategias nacionales de IA, a principios de 2023 solo Brasil consideraba la necesidad de revisar las normas de derechos de autor para incluir excepciones que favorezcan la investigación basada en datos.[64]

Por lo tanto, cuando se trata de obras protegidas por el derecho de autor, es decir, obras que no se encuentran en el dominio público, salvo que estas contemplen una licencia que permita dicho uso, la incorporación de obras sin previa autorización en los *datasets* de entrenamiento de sistemas de IA constituye una violación a los derechos de autor del titular.

Cuando se trata del sistema de Copyright la situación es distinta. La excepción del *fair use*, si bien aún no existe una norma explícita o una resolución judicial que así lo plantee, podría contemplar el uso de obras protegidas para los fines del entrenamiento. Así lo entienden Lemley y Casey (2020) quienes plantean que "los sistemas de machine learning generalmente deberían poder usar bases de datos para el entrenamiento, independientemente de que el contenido de esa base de datos esté protegido por Copyright o no"[65], porque ese uso "suele ser transformador, tal como se entiende ese término en la legislación sobre Copyright, porque, aunque no cambia el trabajo subyacente, cambia el propósito para el que se utiliza el trabajo".[66] Los citados autores manifiestan que en muchos de los casos que se están disputando actualmente los demandados están utilizando elementos de las obras que el propio derecho no protege, como ser las ideas, los hechos o la estructura lingüística. Por ello, Lemley y Casey proponen que el análisis de *fair use* de los datos de entrenamiento contemple lo que denominan el principio de *'fair learning'* –aprendizaje justo–, bajo el cual se analice si el propósito del uso es obtener o incorporar elementos

64 Patricia Díaz Charquero y Viviana Rangel, "Políticas de Inteligencia Artificial y Derechos de Autor en América Latina", *Alianza de La Sociedad Civil Latinoamericana Para El Acceso Justo Al Conocimiento*, febrero de 2023, https://accesoalconocimiento.lat/wp-content/uploads/2023/08/Informe-ia-y-da.pdf

65 Mark A. Lemley y Bryan Casey, "Fair Learning", *Texas Law Review* 99, n.o 4 (30 de enero de 2020): 743-85, https://doi.org/10.2139/ssrn.3528447. La traducción es propia.

66 Lemley y Casey (2020). La traducción es propia.

de una obra sujetos a derechos de autor o acceder, aprender y usar las partes no protegidas de la obra. En el último caso el uso debería ser considerado justo según el primer factor del *fair use* –el propósito y carácter del uso–. En gran parte esta propuesta se funda en los desafíos legales y económicos que plantea la alternativa, es decir, que las compañías desarrolladoras de los sistemas de IA deban obtener licencias de miles, sino millones, de titulares para poder entrenar sus algoritmos. Ahora bien, el *fair learning* no debería ser irrestricto, sino que debería considerar el propósito del sistema y los resultados que genera. Por ejemplo, cuando lo que el modelo de IA pretende obtener de la obra para su entrenamiento tiene que ver con la expresión, entonces los autores plantean que el análisis de *fair use* debería ser más riguroso.

DETERMINACIÓN DE LA AUTORÍA DEL *OUTPUT* GENERADO POR IA GENERATIVA SEGÚN EL MARCO LEGAL VIGENTE

En lo que respecta al *output,* en los países con tradición civil el mayor impedimento para concebir a la IA como autora o creadora de obras deriva de la propia ley. Por ejemplo el artículo 12 de la Ley Federal del Derecho de Autor de México establece que "[a]utor es la persona física que ha creado una obra literaria y artística"; de manera similar la Ley N° 1.328 de Derechos de Autor y Derechos Conexos de Paraguay establece en su artículo 2.1 que se concibe autor a la "persona física que realiza la creación intelectual"; el Decreto Legislativo N° 822 sobre el Derecho de Autor de Perú también en su artículo 2.1 define al autor como la "persona natural que realiza la creación intelectual". Esta identificación del autor con la persona física o natural se replica también en las legislaciones de Brasil,[67] Bolivia,[68] Chile,[69] Colombia[70] y Ecuador,[71] por nombrar algunas. Hay otras jurisdicciones que nada dicen al respecto, mas sus ordenamientos jurídicos no permiten reconocer como autor a una entidad diferente de una persona humana, como es el caso de Argentina.

Esto significa que, al menos en los territorios mencionados, no hay lugar para considerar a la IA como autora hasta tanto cambie la legislación que regula la materia. Por lo tanto, en nuestra región determinar si se trata de contenido

67 Ley N° 9.610 de Derecho de Autor y Derechos Conexos del 19 de febrero de 1998. Artículo 11.
68 Ley N° 1322 de Derecho de Autor del 13 de abril de 1992. Artículo 8.
69 Ley N° 17336 de Propiedad Intelectual del 28 de agosto de 1970. Artículo 5.
70 Ley 23 sobre Derechos de Autor del 28 de enero de 1982. Artículo 8.
71 Ley No. 83. RO/ 320 Ley de Propiedad Intelectual del 19 de Mayo de 1998. Artículo 7.

'generado' por IA o contenido 'asistido' por IA resulta crucial a la hora de determinar si dicho material se encuentra o no protegido por derecho de autor. Según el estado actual de la norma, en el primer caso –contenido generado por IA– estaríamos probablemente ante una obra que se encuentre desde su origen en el dominio público, siendo que no queda amparada por la legislación en materia de derecho de autor. Ahora bien, en lo que respecta a las obras asistidas por IA deberá determinarse en qué medida se utilizó la IA, cuál fue el aporte del agente humano que la utilizó y cuánto de su impronta fue volcada en la obra. Así y todo, podría incluso plantearse una incógnita adicional respecto a quién corresponde asignar la titularidad: ¿al usuario o al desarrollador del programa utilizado?

Concluyendo, en principio, la normativa vigente en Latinoamérica no permite el uso sin autorización de obras para el entrenamiento de sistemas de IA, incluso aquellas que se pudieran obtener mediante la utilización de técnicas de *web scraping*, como así tampoco admite la posibilidad de reconocer autoría sobre las obras generadas por IA. Esto refleja las limitaciones de un marco legal concebido para responder a las realidades de una época pasada que no satisface las exigencias de la nueva era tecnológica.

VI. EL FUTURO DEL DERECHO DE AUTOR EN LA ERA DE LA IA

Como se mencionó anteriormente, el auge de la IA generativa ha suscitado importantes interrogantes en el ámbito del derecho de autor. ¿Es el uso de obras sin licencia para entrenar estos sistemas un 'uso justo' o constituye una infracción? ¿De qué manera y en qué circunstancias deben ser compensados los autores? ¿Los productos generados por IA pueden considerarse obras originales o derivadas, y deberían contar con protección de derechos de autor? ¿Cómo afectaría al dominio público otorgar derechos de autor a obras generadas por estos sistemas? ¿Es necesario distinguir entre creaciones totalmente autónomas y aquellas asistidas por IA?

Las demandas del mercado y de las partes interesadas dejan en evidencia que una adecuación normativa que contemple las necesidades de la realidad tecnológica actual, protegiendo a los autores y sus obras originales, es ineludible. Frente a esto resulta crucial reflexionar sobre las distintas opciones y caminos que se vislumbran para el derecho de autor en América Latina. A continuación se analizan tres posibles escenarios para el futuro de la regulación de las creaciones en la era de la IA:

- Escenario Uno - El primer escenario propone modernizar la normativa a fin de que esta contemple las particularidades de la relación entre las obras y los sistemas de IA. En concreto este escenario requiere que cada jurisdicción tome una postura respecto al tratamiento del *input* y el *output*, por ejemplo modificando las legislaciones vigentes para contemplar el uso de obras como *input* –determinando si el uso de obras protegidas en el entrenamiento de IA requiere compensación o no– y aportando claridad jurídica respecto al tratamiento de las creaciones derivadas de la aplicación de IA, en particular IA generativa. Las excepciones, limitaciones o cambios que son necesarios introducir para que el derecho de autor continental conciba las particularidades de esta tecnología requieren replantear las bases de este derecho a fin de flexibilizar las barreras legales impuestas por la letra actual de la ley.
- Escenario Dos - Alternativamente se podría considerar si existen otras áreas del derecho, distintas del derecho de autor, que puedan proporcionar una regulación más eficaz y adecuada para gestionar las implicancias del desarrollo de los sistemas de IA. En un artículo publicado por el Yale Law Journal,[72] Micaela Mantegna explica que modificar el derecho de autor para cubrir las necesidades que plantea la IA –Escenario Uno– puede ser perjudicial para la sociedad en general, afectando la cultura y la innovación. Siendo que las consecuencias de la IA generativa no se limitan al derecho de autor, Mantegna plantea la necesidad de políticas holísticas que permitan considerar el problema desde una perspectiva más amplia, aplicando una variedad de marcos jurídicos –derechos humanos, derecho laboral, protección de datos personales y derecho del consumidor–.
- Escenario Tres - Una tercera opción podría ser plantear la creación de una categoría jurídica *sui generis* –tal como algunas jurisdicciones han hecho en materia de bases de datos o conocimientos tradicionales– que regule el punto de conexión entre las creaciones y la IA, es decir, el *input* y el *output*.

Resulta interesante analizar cuáles serían las implicancias de cada uno de estos escenarios y cómo podría verse el futuro del derecho de autor ante la era de la IA.

72 Micaela Mantegna, "ARTificial: Why Copyright Is Not the Right Policy Tool to Deal With Generative AI," The Yale Law Journal 133 (April 22, 2024): 1126, https://www.yalelawjournal.org/forum/artificial-why-copyright-is-not-the-right-policy-tool-to-deal-with-generative-ai.

ESCENARIO UNO: MODERNIZACIÓN DEL DERECHO DE AUTOR

El primer escenario propone actualizar la normativa vigente, lo cual requiere adoptar una postura ante las preguntas que surgen en relación con esta tecnología, tanto en lo que respecta al *input* como al *output*, lo cual se traduciría en cambios sustanciales como la incorporación de nuevas excepciones al uso de obras protegidas y la redefinición de conceptos centrales del derecho de autor continental como los de originalidad, creatividad y autoría.

Por un lado, en lo que respecta al *input*, cada territorio deberá definir si considera el uso de obras protegidas para los fines del entrenamiento de algoritmos de IA una violación o no a los derechos de autor de sus titulares. Se presenta una decisión que a primera vista parece ser un juego de suma cero y que enfrenta a los autores contra las empresas desarrolladoras de estos sistemas de IA. Sin embargo, desde el plano legislativo se podrían plantear soluciones intermedias en dónde, de manera similar a las licencias obligatorias, se incorpore una limitación a los derechos de autor de los titulares, debiendo estos tolerar su uso para el entrenamiento de los algoritmos a cambio del pago de una contraprestación. Sin perjuicio de las decisiones que adopte cada legislación, será de suma importancia establecer espacios de diálogo entre los países de América Latina y procurar tener una mirada regional sobre estos temas.

Por otro lado, en lo que respecta al *output*, los avances de la IA generativa y su creciente aplicación en el ámbito creativo presentan un desafío legal que exige ser tratado. Otorgar protección a las obras generadas por IA bajo el régimen de derecho de autor requiere que se reconsidere la naturaleza de conceptos que forman las bases de este derecho como lo conocemos, conceptos como creatividad y originalidad, hasta incluso el propio concepto de autoría. A pesar de que la IA es ya una herramienta omnipresente, seguimos enfrentándonos a grandes interrogantes sobre su capacidad para asumir el rol de autora, tanto en el presente como en el futuro. Es crucial explorar los principios esenciales de la autoría, como la creatividad y la originalidad, para entender el verdadero alcance de su potencial en este terreno. A continuación se analizarán brevemente estos términos:

A. **Creatividad:** desde el punto de vista etimológico, el término 'creatividad' proviene del latín *creare* que puede traducirse como "engendrar", "producir" o "crear". La aplicación de herramientas digitales en el proceso creativo no es una novedad, hace años estas se han convertido en grandes aliadas de diseñadores y creativos en general. Sin embargo, el uso de IA generativa no es asimilable a los softwares tradicionales de diseño ya que en ese tipo de

programas el usuario al menos debe aplicar su propio criterio en el proceso creativo a fin de obtener un resultado que refleje su estilo y personalidad. La diferencia que introduce la IA generativa tiene que ver con que la intervención del usuario es cada vez menor, por lo tanto también lo es su impronta sobre el resultado final. No es necesario que se le proporcione instrucciones elaboradas al programa para que este arroje un resultado que, sin conocer el contexto,[73] puede ser considerado creativo. Si concebimos que el *output* generado por los sistemas de IA generativa puede ser catalogado como creativo, podríamos estar ante el primer paso para contemplar la capacidad de la IA generativa de crear.

Esto nos lleva a cuestionarnos si la conciencia creativa es realmente esencial en el proceso de creación de una obra y a considerar si la IA podría desarrollar dicha conciencia en un futuro cercano.[74] Pensadores como Block[75] han sostenido que la conciencia está arraigada en la biología y que los sistemas sintéticos no pueden tener experiencias subjetivas. Otros, como Shevlin,[76] afirman que no se necesitan cerebros biológicos para la conciencia y que lograremos una IA consciente para finales del siglo. McClelland sostiene que ambas posturas están equivocadas, porque aún no sabemos lo suficiente sobre la conciencia como para emitir un juicio informado sobre

73 En 2022 Jason Allen, un diseñador de videojuegos del estado de Colorado (EE.UU), ganó el primer puesto en la categoría de arte digital de la Feria Estatal de Colorado por su obra *Théâtre D'opéra Spatial*, la cual había sido realizada con el programa Midjourney que permite generar imágenes a partir de texto. Si bien Allen alega haber informado sobre el uso del sistema de IA en la generación de la obra, los jurados sostuvieron que tomaron conocimiento de esa situación luego de haber otorgado el premio una vez que el propio Allen publicó esta información en sus redes sociales. Este caso deja en evidencia las capacidades de estas herramientas y la importancia de revelar su uso, ya que el mismo –por lo general– no puede ser detectado a simple vista.

74 Para analizar la cuestión de la conciencia en la IA, ver Clare College - University of Cambridge, "Will AI Ever Be Conscious?", s. f., https://stories.clare.cam.ac.uk/will-ai-ever-be-conscious/index.html.

75 Talks at Google, "AI Cognition Won't Work for Consciousness | Ned Block | Talks at Google", 14 de junio de 2017. Video, https://www.youtube.com/watch?v=6lHHxcxurhQ

76 Henry Shevlin, "Non-human Consciousness And The Specificity Problem: A Modest Theoretical Proposal", *Mind & Language* 36, n.o 2 (18 de enero de 2021): 297-314, https://doi.org/10.1111/mila.12338

las posibilidades de una IA consciente.[77] Por su parte, Margaret Boden,[78] investigadora inglesa especialista en ciencia cognitiva, entiende que desde una perspectiva psicológica la creatividad es la habilidad de generar formas, ideas o artefactos que sean nuevos, sorprendentes y valiosos. Según esta autora el desafío de los ordenadores –léase la IA– en relación con la creatividad se centra en que no poseen sentido de relevancia, por lo que, si bien pueden combinar ideas o información de manera novedosa, no poseen la capacidad de distinguir si estas o incluso los resultados que arrojan –*output*– son interesantes o relevantes.[79]

B. **Originalidad:** el requisito mínimo para que una obra sea protegida mediante las normas de derecho de autor suele ser la originalidad. Si bien algunas jurisdicciones –como por ejemplo México y Estados Unidos– exigen adicionalmente que la obra sea fijada en un medio de expresión tangible,[80] en líneas generales, no son relevantes a los fines de la existencia de la protección el mérito de la obra, su forma de expresión, su destino,[81] ni cuán novedosa sea. Tampoco lo es su registro, condición que deriva del artículo 5.2 del ya mencionado Convenio de Berna.

La OMPI explica que "[u]na obra original es una obra creada independientemente y no copiada de otra ni del material que forme parte del dominio público".[82] Sin embargo, en América Latina, los tribunales han ido más allá, sosteniendo consistentemente que la originalidad de una obra debe reflejar una individualidad distintiva, que plasme la impronta de su autor de manera clara y evidente.[83]

---

77 Timothy Nguyen, "Jay McClelland | Neural Networks: Artificial and Biological | The Cartesian Cafe with Timothy Nguyen", 2 de octubre de 2024. Video, https://www.youtube.com/watch?v=yQbJNEhgYUw.

78 Margaret A. Boden, "Creativity And ALife", *Artificial Life* 21, n.o 3 (1 de agosto de 2015): 354-65, https://doi.org/10.1162/artl_a_00176.

79 Fundación Telefónica, "Transformación Digital: Margaret Boden | #CitasFTef", 13 de mayo de 2019. Video, 0:01:26, https://www.youtube.com/watch?v=k5VWIvbd9bU.

80 Copyright Act, 1976, 17 U.S.C. §102.

81 Eduardo De La Parra Trujillo, "Nociones básicas sobre el objeto de los derechos de autor", *Revista Del Instituto de La Judicatura Federal* 18 (2004), https://escuelajudicial.cjf.gob.mx/publicaciones/revista/18/r18_4.pdf.

82 "¿Cómo obtener protección por derecho de autor?", WIPO - World Intellectual Property Organization, s. f., https://www.wipo.int/copyright/es/protection.html#:~:text=Para%20quedar%20protegida%20por%20derecho,con%20la%20idea%20de%20base.

83 Tribunal de Justicia de la Comunidad Andina, 31 enero, 2020 Criterios jurísprudenciales sobre la originalidad de una obra y el plagio de la misma, consultado en

Este requisito tiene que ver con la forma de expresión de la obra y no así con la idea de fondo. Requiere un aporte individual y creativo, es decir, producto de un pensamiento independiente, que aporte a la obra algo individual y propio de quien la crea.[84]

Determinar si las producciones de los sistemas de IA generativa son originales o no dependerá en gran medida de cómo estos funcionan y de cuánta injerencia tenga el usuario sobre el *output*. Cuando se interactúa con programas como ChatGPT de la empresa OpenAI, Claude de Anthropic o Gemini de Google –algunos de los *chatbots* de IA más avanzados en el mercado– pareciera a primera vista que estos realmente generan contenido original. Sin embargo, reclamos como los presentados por The New York Times o Concord Music Group, Inc. denotan que estos sistemas aún replican parte de los datos de entrenamiento. Independientemente de si las legislaciones permitieran o no concebir a los sistemas de IA como autores, una de las principales barreras que deberá superar quien pretenda proteger materiales generados con herramientas de IA será demostrar su originalidad.

c. **Autoría:** como se mencionó anteriormente, actualmente existe un impedimento legal para concebir a la IA como autora. Incluso en el sistema de Copyright para considerar a una creación como una 'obra original de autoría' se requiere que esté creada por un ser humano.[85] En los países de tradición civil, esta limitación legal deriva naturalmente de la estructura que estos han concebido para la protección de las obras. El derecho de autor continental reconoce al autor derechos morales que están íntimamente ligados a la persona que crea, son inalienables e irrenunciables, por lo que no se podría concebir que recaigan sobre cabeza de un agente distinto de la persona humana o física que engendra la obra.

¿Quién es entonces el autor de las obras generadas por sistemas de IA? Bajo la legislación vigente en la región podríamos tener tres posibles respuestas: la persona que programó el modelo, el usuario que ha ingresado las instrucciones al sistema para obtener ese resultado, o simplemente nadie. La primera opción quizás podría tener sentido si habláramos de sistemas de IA tradicional, sin embargo, como se ha comentado, los sistemas de IA generativa funcionan de manera diferente, pudiendo mediante su

https://www.tribunalandino.org.ec/index.php/2020/01/31/criterios-jurisprudenciales-sobre-la-originalidad-de-una-obra-y-el-plagio-de-la-misma/

84 Delia Lipszyc, *Derecho de Autor y Derechos Conexos* (UNESCO-CERLALC-ZAVALÍA, 1993).

85 U.S. Copyright Office, Compendium of U.S. Copyright Office Practices (3d ed. 2021). Chapter 300, página 21.

entrenamiento y aprendizaje propio alejarse de su programación inicial. La segunda opción quizás pareciera ser la más acertada, sin embargo hay usuarios que alegan no haber realizado un aporte lo suficientemente relevante como para ser considerados autores. Justamente uno de los desafíos que presenta la utilización de IA en el proceso creativo es la dificultad de determinar en qué proporción el material ha sido creado gracias al aporte del autor, más allá de si el resultado cumple con los requisitos mínimos para ser considerado obra. Sin perjuicio de todo lo anterior, si no se reconoce originalidad a los materiales generados por sistemas de IA y no existe tampoco un claro autor a quien asignar los derechos, pareciera que la opción más viable es la tercera, la cual implicaría que todos estos materiales recaigan de manera automática en el dominio público, permitiendo que la producción generada por IA sea accesible para todos, ampliando así el conocimiento y los recursos culturales sin crear restricciones artificiales.

Si se pretende reconocer que los sistemas de IA generativa efectivamente crean, que sus resultados son originales y que, por lo tanto, se les puede adjudicar autoría sobre sus creaciones, entonces se requiere innegablemente un ajuste legislativo. En parte, para ello es necesario dejar de lado la concepción antropocéntrica de la autoría que solamente concibe a la persona física o humana como autor, sin embargo esto no es suficiente. Estos cambios derivan en otras preguntas que si bien exceden el objeto del presente es necesario considerarlas. Si contemplamos la posibilidad de una IA autora, ¿cómo se pretende asignarle derechos? ¿Debemos considerar entonces a la IA como una persona jurídica? ¿Es necesaria una nueva categoría jurídica que conciba a estos sistemas como sujetos de derecho? Las respuestas una vez más no son sencillas y requieren de un estudio profundo que contemple no solamente cómo estos cambios podrían afectar el futuro de la IA sino de la sociedad en su conjunto.

Los cambios necesarios parecieran ser demasiado drásticos e implicarían una completa modificación de las bases del derecho de autor como lo conocemos. En lugar de modificar directamente la forma en que estos términos están previstos en nuestras legislaciones, una alternativa más viable sería ajustar cómo los interpretamos, realizando pequeños cambios que permitan la coexistencia de la creación humana y la creación artificial, concibiendo la posibilidad de una creatividad algorítmica. Si entendemos que el algoritmo tiene la capacidad de generar resultados creativos –crear–, podríamos incluso contemplar la posibilidad de una originalidad colaborativa entre el usuario y el modelo de IA, en donde se reconozca tanto el peso de los aportes del usuario que provee las instrucciones

como del sistema que las interpreta y genera resultados basados en su propio aprendizaje.

ESCENARIO DOS: REGULACIÓN A TRAVÉS DE OTRAS RAMAS DEL DERECHO

Alternativamente deberíamos analizar la posibilidad de que el derecho de autor no sea la herramienta indicada para abarcar los aspectos legales relacionados al *input* y *output* de los modelos de IA generativa. Esto ha sido planteado por algunos de los autores ya mencionados en el presente, como Lemley y Casey (2020) quienes manifiestan que debido a la cantidad de obras involucradas en el entrenamiento de los algoritmos, cuando efectivamente ocurran abusos a los derechos de autor de los titulares, el Copyright puede no ser la herramienta adecuada para regular esa situación; y Mantegna (2024) quien propone aplicar una variedad de marcos jurídicos a la relación de la IA con las obras, sin estrictamente limitarlo al derecho de autor.

En lo que respecta al *input*, como ya se ha mencionado, el principal desafío se encuentra en que los *datasets* de entrenamiento pueden incluir miles de obras de una gran variedad de titulares, por lo que licenciar individualmente todo el material contenido en estos no es una opción factible para los desarrolladores de sistemas de IA, ya que ello implicaría altos costos administrativos como así también demoras en los procesos. La falta de claridad legal sobre cómo gestionar los derechos de los autores ante estos usos afecta de manera directa a los titulares de las obras, siendo que ante la inexistencia de una regulación específica en la materia, por el momento, se están viendo forzados a tolerar el uso de sus creaciones sin autorización y sin compensación alguna. Ante la falta de mejores herramientas para defender y gestionar sus activos creativos, y la aparente imposibilidad material de exigir que sus obras sean sustraídas de las bases de datos de entrenamiento, estos recurren a reclamos fundados en violaciones a sus derechos de autor, cuando es probable que en muchas ocasiones este no sea el caso. La 'sobreinclusión' en la aplicación de los derechos de autor es uno de los problemas a los que hacen referencia los citados Lemley y Casey, identificando muy acertadamente que "[l]a legislación sobre derechos de autor otorga a los propietarios el control sobre algunas partes de su obra, pero no sobre otras. Sin embargo, cuando los usuarios no pueden separar las partes que se pueden proteger de las que no, el control sobre una parte puede convertirse fácilmente en el control sobre todo. Esto puede ocurrir por accidente, pero también puede ocurrir deliberadamente. Los propietarios de

derechos de autor utilizan habitualmente la legislación como herramienta para impedir la competencia disruptiva que amenaza sus mercados actuales".[86]

El uso de obras protegidas para el entrenamiento de algoritmos de IA no plantea un problema simplemente por la falta de autorización, sino que a ello se le suma la falta de compensación. Esto no solamente perjudica económicamente a los autores privándolos de una fuente adicional de ingresos, sino que podría considerarse como '*free riding*'[87] por parte de las empresas de IA. El Tribunal de Justicia de la Unión Europea ha entendido en cuestiones de marcas –las cuales podrían extrapolarse al ámbito comercial del derecho de autor– que el *free riding* "no se vincula al perjuicio sufrido por la marca, sino a la ventaja obtenida por el tercero del uso del signo idéntico o similar",[88] por lo que el foco del análisis debería situarse sobre el beneficio de estas empresas en el uso de las obras y no en si dicho uso afecta a las obras, o subsidiariamente, a sus autores. En resumidas cuentas, si utilizan propiedad de terceros para desarrollar sus sistemas, actividad que tiene un evidente fin económico, y no compensan a los titulares de esas creaciones por dicho uso, estamos ante un comportamiento que desde la perspectiva del derecho de competencia podría hasta ser considerado como un aprovechamiento indebido del esfuerzo ajeno, conducta que suele ser considerada ilícita.

Asimismo, hay gran preocupación en la comunidad creativa con respecto al impacto que los materiales generados con estas herramientas puedan tener en sus ingresos, siendo que estos no solamente competirán con la IA por oportunidades laborales sino que además sus obras competirán con el output de los sistemas de IA. Si bien es poco probable que la IA reemplace a los creadores, no se debe aminorar el impacto que esta tecnología puede tener en el mercado final. La incorporación de la IA generativa al mercado creativo ha ampliado exponencialmente la oferta de 'mano de obra', aunque artificial, y de obras sintéticas, lo cual efectivamente ha alterado las oportunidades económicas de los autores. Si bien es cierto que las creaciones de la IA aún suelen tener falencias y es difícil imaginar que estas puedan equipararse a las creaciones humanas, el mercado compartido puede tener distintas fuentes de originación, entre ellas la IA generativa. En este sentido, no es necesario que

86 Lemley y Casey (2020). La traducción es propia.

87 Término utilizado en la economía para describir la conducta de un agente que usa o recibe los beneficios del uso de un bien o servicio sin haber pagado por el.

88 Aurea Suñol, "Aprovechamiento Indebido de la Reputación de una Marca y Free-riding", Almacén de Derecho, 1 de septiembre de 2015, https://almacendederecho.org/aprovechamiento-indebido-de-la-reputacion-de-una-marca-y-free-riding

vaticinemos un futuro en donde la IA reemplace el trabajo humano creativo para concebir a esta tecnología como un competidor real que tiene la capacidad para ocupar parte de ese mercado. No cabe duda de que, si bien la IA puede ser una herramienta valiosa para potenciar la creatividad, también plantea un desafío subyacente estrechamente ligado a la competencia.

Un reporte publicado a fines de 2023 por la Comisión Federal de Comercio de Estados Unidos enfocado en el impacto de la IA generativa en la economía creativa ya alertaba sobre estos riesgos previendo que "[c]uando se utiliza la IA en las profesiones creativas pueden surgir diversas preocupaciones en materia de competencia y protección del consumidor. Conductas como entrenar una herramienta de IA para que utilice expresiones protegidas sin el consentimiento del creador o vender el producto generado por esa herramienta de IA, incluso imitando el estilo de escritura, la interpretación vocal o instrumental o la imagen del creador, puede constituir un método de competencia desleal o una práctica desleal o engañosa".[89]

Desde una perspectiva de derecho de la competencia, el uso no autorizado y no compensado de obras para desarrollar programas que luego competirán con estas y sus autores, disminuyendo a su vez muy probablemente el valor de dichas obras existentes o las futuras, podría considerarse un ilícito. Sin embargo, ya sea que apliquemos el derecho de autor o el derecho de la competencia, el problema de la implementación –*enforcement*– en relación a obras determinadas o titulares de derecho individualmente identificados subsiste. La diferencia reside en que los organismos gubernamentales de defensa de la competencia suelen tener la facultad de sancionar conductas anticompetitivas, tales como abusos de posición dominante y hasta carteles, por lo que podría resultar ser una herramienta más efectiva para hacer valer los derechos de los titulares afectados.

Otro de los marcos legales que podrían aplicarse en particular a los *datasets* de entrenamiento es el que regula la protección de datos personales. Si bien este no protege las obras en sí mismas, sí podría ser utilizado por los titulares de estas como una herramienta legal adicional para hacer valer sus derechos. Parte de los datos que conforman el *input* serán indefectiblemente datos personales –como por ejemplo los datos del autor de una obra–, sobre todo a la luz de

89 US Federal Trade Commission, "Generative Artificial Intelligence and the Creative Economy Staff Report: Perspectives and Takeaways", Federal Trade Commission, diciembre de 2023, https://www.ftc.gov/system/files/ftc_gov/pdf/12-15-2023AICEStaffReport.pdf. La traducción es propia.

normativas que prevén definiciones amplias de este término, como la mexicana[90] o la argentina.[91] El tratamiento de estos datos suele requerir el consentimiento expreso del titular, por lo que su uso no autorizado podría implicar una violación a este tipo de normas. Sin embargo, algunas legislaciones contemplan a su vez excepciones que prevén técnicas como la mencionada *web o data scraping*[92] por lo que esta regulación pierde valor estratégico a la hora de ejercer una defensa por parte de los titulares de obras protegidas.

ESCENARIO TRES: NUEVA CATEGORÍA *SUI GENERIS*

La IA generativa no es el primer avance tecnológico que ha desafiado al derecho. En las últimas décadas varias normas han debido actualizarse para contemplar el nuevo paradigma planteado por la integración del mundo físico y digital. Sin embargo esto no siempre es suficiente. Los escenarios desarrollados anteriormente, si bien podrían ser opciones viables para hacer frente a los desafíos que se han mencionado en el presente, se basan en estructuras que no fueron pensadas para los retos que presenta la IA. El Escenario Uno propone modificaciones que podrían tener implicancias legales que no llegamos a prever, lo cual puede resultar problemático y desventajoso. Por su parte, las diversas áreas del derecho mencionadas en el Escenario Dos no proporcionan herramientas ideales para gestionar la relación entre las obras y el continuo desarrollo de los sistemas de IA.

Se podría alegar que las particularidades que plantea esta tecnología, su dependencia de datos –gran parte de los cuales resultan ser obras protegidas– y su creciente aplicación en el ámbito creativo, se verían beneficiados de un marco legal *sui generis* independiente del derecho de autor que contemple las características del *input* y el *output*. En lo que respecta al *input*, como se ha mencionado reiteradas veces, el principal problema a resolver se centra en la dificultad material de obtener las licencias correspondientes para el uso

90 El artículo 3 de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares define a los datos personales cmo "cualquier información concerniente a una persona física identificada o identificable".

91 El artículo 2 de la Ley 25.326 de Protección de los Datos Personales (LPDP) define a los datos personales como "información de cualquier tipo referida a personas físicas o de existencia ideal determinadas o determinables".

92 Por ejemplo el artículo 5.2 de la LPDP argentina establece que no es necesario el consentimiento cuando "los datos se obtengan de fuentes de acceso público irrestricto", como Internet.

de obras protegidas en el entrenamiento de los algoritmos de IA. Esta nueva categoría propuesta podría implementar un sistema de retribuciones para las obras utilizadas a tales fines. No implicaría técnicamente la incorporación de una nueva excepción al derecho de autor sino más bien una nueva estructura de monetización que permita a los autores redituar de su trabajo en la era de la IA.

En lo que respecta al *output*, una regulación *sui generis* permitiría establecer un marco regulatorio para las obras generadas por IA o con asistencia de sistemas de IA, identificando expresamente quiénes son los titulares de dichos materiales y cuáles son los derechos que se les reconocen. Mediante la incorporación de reglas claras, la solución propuesta en este Escenario Tres aportaría a nivelar el campo de juego entre los artistas humanos y la tecnología en cuestión. Por ejemplo, considerando el ritmo al cual la IA generativa puede crear contenido, se podría condicionar la protección de obras creadas mediante la aplicación de sistemas de IA a un registro. De esta manera solamente se otorgaría protección a aquellas obras que su autor o la persona involucrada en el proceso de creación considere meritorio como para solicitar dicho registro y abonar el arancel establecido a tal fin. Sujetar estas creaciones a un proceso de registro permitiría la recopilación de información sobre la obra –como podría ser respecto al sistema utilizado y hasta los *prompts* o indicaciones que permitieron al usuario llegar a dichos resultados– y proporcionaría una oportunidad para su revisión.

En cuanto a la duración de la protección, un plazo más acotado que el que establece el derecho de autor permitiría satisfacer los diferentes intereses involucrados. No hace falta ahondar en el rol que juega el reconocimiento de una exclusividad en el fomento del desarrollo de nuevas obras, asimismo plazos más cortos permitirían a su vez que estas obras aporten en el corto o mediano plazo –según defina cada jurisdicción– a la sociedad, enriqueciendo el patrimonio cultural. Al igual que sucede en materia de derecho de autor se podrían prever distintos plazos de protección según el tipo de obra de que se trate, considerando el esfuerzo que requiere la producción del material o el nivel de intervención humana en el proceso de creación.

La IA generativa tiene la capacidad de generar resultados que muchas veces son difíciles de diferenciar de aquellos realizados por personas naturales, esto representa un desafío cada vez mayor para los artistas para quienes la IA se ha convertido en un real competidor, poniendo en jaque su capacidad de generar ingresos. Ante esto, las grandes plataformas de contenido como Instagram y

Facebook de Meta[93] o TikTok[94] han aplicado políticas respecto al etiquetado de contenido generado por IA a fin de fomentar la transparencia. Una regulación como la que plantea el Escenario Tres podría exigir el etiquetado del contenido creado mediante la aplicación de sistemas de IA, homogeneizando así el tratamiento de estos materiales, dejando de ser su identificación discrecional.

Los mencionados son parámetros generales que permitirían una mejor convivencia entre los artistas, los sistemas de IA y sus desarrolladores. Una regulación de ese estilo no sería ajena al contexto socioeconómico de cada jurisdicción y es posible que sea necesario que existan variaciones de país en país. Sin embargo, tal como se ha mencionado anteriormente, si bien este tipo de regulaciones pueden tener éxito en jurisdicciones aisladas, se deben pensar de manera regional y ser acompañadas idealmente de una estrategia interjurisdiccional que permita la integración en estos temas, al menos, de los países de la región.

Por último, considerando que no hemos evidenciado aún el potencial completo de esta tecnología, sería importante que una normativa como la que contempla el Escenario Tres tenga en cuenta el dinamismo que ha demostrado la IA y prevea la posibilidad de realizar revisiones periódicas respecto a la eficacia de su aplicación.[95] Tanto los avances de la propia tecnología como las novedades legislativas o jurisprudenciales que surjan, podrían impactar en la eficacia de disposiciones como las aquí planteadas.

93 "Cómo se identifica el contenido de IA en los productos de Meta", Meta, s. f., https://www.meta.com/es-es/help/artificial-intelligence/how-ai-generated-content-is-identified-and-labeled-on-meta/?srsltid=AfmBOopTp7Yy-ffDzwwoKbw51AMkz8oUGg6rzsKQl0YvM2bZsqFv3Un_

94 "Acerca del Contenido Generado Por IA", Centro de Ayuda de TikTok, s. f., https://support.tiktok.com/es/using-tiktok/creating-videos/ai-generated-content.

95 La Directiva Europea 96/9/CE que introdujo un régimen legal *sui generis* en materia de protección de bases de datos incluye en su artículo 16.3 la obligación de presentar informes cada tres años antes el Parlamento Europeo, el Consejo de la Unión Europea y el Comité Económico y Social sobre su aplicación y los efectos de la aplicación del nuevo régimen planteado, acompañado de propuestas destinadas a actualizar la norma en base a la evolución del objeto de protección –las bases de datos–.

VI. PROYECTOS DE REGULACIÓN EN LA REGIÓN

Varios países de nuestra región han publicado o sancionado instrumentos legales que plantean directrices, principios o buenas prácticas en materia de IA, mas no existe en América Latina aún una regulación concreta para esta tecnología. Esto puede cambiar en cualquier momento, siendo que durante el último año han comenzado a surgir proyectos de ley en algunos territorios que, siguiendo los pasos de Europa, pretenden justamente regular la IA, podemos mencionar entre ellos a Chile, Perú, Colombia, Argentina y Ecuador.

Entre los proyectos presentados se destaca el de Ecuador denominado "Proyecto de Ley Orgánica de Regulación y Promoción de la Inteligencia Artificial en Ecuador"[96] el cual además de plantear una estructura de análisis de riesgos similar a la Ley de Inteligencia Artificial Europea que entró en vigor en 2024 –lo cual se puede apreciar también en los proyectos de Argentina y Chile–, incluye un capítulo destinado en particular a los derechos de autor y derechos conexos.

En ese capítulo el mencionado proyecto incluye previsiones relacionadas tanto al *input* como al *output*. En lo que respecta al primero, se establece que el uso de obras protegidas por derechos de autor y derechos conexos para el entrenamiento de sistemas de IA estará permitido sin requerir autorización previa cuando esta actividad tenga fines de investigación científica no comercial por parte de universidades, centros de investigación e instituciones culturales. Esta previsión establece requisitos mínimos como por ejemplo incluir el reconocimiento de la fuente y paternidad de las obras siempre que sea posible y que dicho uso se realice en la medida justificada por el objetivo no lucrativo perseguido. En cambio, cuando se trate de usos comerciales o a gran escala de obras protegidas como parte de los *datasets* de entrenamiento se requerirá de licencias individualmente negociadas o licencias colectivas mediante la intervención de entidades de gestión legalmente autorizadas.

En relación con el *output* o las obras en cuya generación estuvo involucrado un sistema de IA, el mencionado proyecto distingue entre obras generadas con asistencia de IA y aquellas que fueron generadas de manera autónoma por IA. En el caso de las primeras, siempre y cuando sean originales y exista una contribución creativa humana sustancial, se les reconoce protección por

96 Disponible para consulta en https://observatorioecuadordigital.mintel.gob.ec/wp-content/uploads/2024/08/PP_-_Proyecto-de-ley-450889-nunez_c.pdf

derecho de autor identificando al creador humano como autor y titular originario. En cambio, las obras 'autogeneradas' por IA, en principio, formarán parte del dominio público. Sin perjuicio de ello, se reconoce acciones de competencia desleal a la persona que haya creado, entrenado o sea propietaria del sistema de IA que generó la obra a fin de prevenir usos no autorizados que le causen un perjuicio económico injustificado. El mismo proyecto establece en otro artículo que "las obras generadas de manera autónoma por IA que sean intrínsecamente incapaces de satisfacer el requisito de originalidad serán consideradas parte del dominio público", lo cual denota la relevancia de conceptos como el de 'originalidad' discutido anteriormente.

En Colombia, el primer periodo de la legislatura 2024-2025 cerró el año con doce proyectos de ley en trámite que proponían regulaciones relacionadas a IA en una gran variedad de áreas.[97] Entre ellos solo uno trata temas relacionados a los aquí desarrollados. El Proyecto de Ley 293 de 2024 [PL 293-2024(S)][98] presentado ante el Senado concentra en 8 artículos "lineamientos para el entrenamiento de modelos o sistemas de inteligencia artificial (IA) y [define] la gestión colectiva obligatoria de algunas formas de uso de obras protegidas por derecho de autor". Este breve proyecto de ley establece que es necesaria una licencia para el uso de obras protegidas por derecho de autor dentro de los datos de entrenamiento, la cual debe ser procurada por el titular del sistema previo al inicio del entrenamiento de modelo de IA. Se determina que dicho uso será remunerado, debiendo las obras utilizadas en el entrenamiento ser reportadas por el responsable del sistema a las sociedades de gestión de derecho de autor y derechos conexos correspondiente. Adicionalmente, incluye una excepción relacionada a la minería de textos y datos con fines de investigación científica.

97 Prensa Jurídica, "Así Van los Doce Proyectos de Ley Sobre Inteligencia Artificial (IA) En Colombia", 19 de diciembre de 2024, https://www.prensajuridica.com/details/item/32613-as%C3%AD-van-los-doce-proyectos-de-ley-sobre-inteligencia-artificial-ia-en-colombia.html

98 Disponible para consulta en https://www.prensajuridica.com/A-PI-EDICION-30-2024/PL%20293-2024(S)-IA%20DERECHOS%20DE%20AUTOR.%20Ponencia%201er%20debate%20Com6%20Senado.pdf

VIII. CONCLUSIÓN

Este artículo ha abordado dos aspectos clave en la relación entre la inteligencia artificial y los derechos de autor: la autoría del *output* y el uso de obras protegidas como *input* en el entrenamiento de los sistemas de IA. El problema del *output* es relativamente sencillo de analizar: en los últimos tiempos, se ha concluido en numerosas ocasiones que la intervención humana directa es esencial para la autoría. Los contenidos generados de forma totalmente autónoma por un sistema de IA no deberían recibir protección por derechos de autor, ya que otorgarles exclusividad afectaría los objetivos de este derecho y limitaría el acceso al dominio público sin aportar beneficios claros a la creatividad.[99] Aunque algunos sistemas jurídicos han comenzado a permitir la protección de ciertas obras generadas por IA,[100] en general, la legislación sigue exigiendo autoría humana para reconocer derechos de autor. En cambio, el *input* plantea un desafío más complejo, ya que muchos sistemas de IA se entrenan con material protegido por derecho de autor sin permiso de ni compensación para los creadores. Como se ha comentado, este uso no autorizado ha generado un intenso debate sobre la necesidad de nuevas regulaciones que equilibren el desarrollo tecnológico con la protección de los derechos de los titulares de contenido.

Las reflexiones presentadas en este artículo evidencian las limitaciones de la legislación sobre derechos de autor en su forma actual para afrontar los grandes desafíos que plantea la IA. Se han analizado algunos de los casos actuales pero aún no se sabe si estos casos darán las respuestas que se necesitan para abordar el potencial disruptivo de la IA generativa. ¿Lograrán proteger los derechos de los titulares de las obras protegidas, mientras fomentan la evolución de esta herramienta de tal manera que enriquezca y potencie la creatividad humana, o estamos depositando expectativas desmesuradas en un sistema jurídico que no fue diseñado para esta era tecnológica? Frente a estos desafíos tan complejos, cabe preguntarse si la estructura del derecho de autor demostrará ser lo suficientemente robusta para adaptarse y evolucionar, o si sus fundamentos quedarán expuestos como insuficientes y obsoletos.

¿Cuál es la solución? ¿Cómo deberíamos equilibrar la necesidad de compensar a los creadores de contenido por su trabajo, mientras fomentamos el de-

99 WIPO Conversation on AI and IP, 10th session, https://www.wipo.int/meetings/en/details.jsp?meeting_id=84809

100 Edward Chatterton y Wilson Lung, "Legal protections for AI-generated work in China", The Global Legal Post, 16 de febrero de 2024, https://www.globallegalpost.com/news/legal-protections-for-ai-generated-work-in-china-45229593

sarrollo de nuevas tecnologías? Este es el debate complejo que nos interpela. Las discusiones sobre cómo equilibrar la compensación a los creadores de contenido con el fomento de nuevas tecnologías, como la IA, plantea cuestiones fundamentales. Como hemos señalado, las empresas que desarrollan sistemas de IA podrían copiar material protegido por derechos de autor y reutilizarlo indefinidamente, lo que genera preocupaciones no sólo sobre la ética y justicia de esta práctica sino también sobre la naturaleza anticompetitiva de esta conducta. Este tipo de uso de las obras puede ser considerado injusto si no se remunera adecuadamente a los creadores.[101]

Como conclusión, nuestra reflexión final tomará como referencia a la industria musical para examinar su respuesta ante el advenimiento de la IA. Este es un sector en el que hemos acumulado una experiencia especialmente rica y significativa a lo largo de los años, por lo que analizaremos cómo ha afrontado en el pasado tecnologías disruptivas y desafiantes. Creemos que este análisis puede ofrecer una perspectiva valiosa y servir de ejemplo para orientar a otras industrias en este nuevo escenario.

Los modelos de IA generativa están creando canciones que no siempre son fáciles de distinguir de las composiciones humanas. Como sabemos, estos modelos se entrenan con grandes conjuntos de datos no licenciados obtenidos de Internet y plataformas de contenidos musicales, lo que supuestamente infringe los derechos de autor de millones de artistas. Sin embargo, la responsabilidad legal sigue siendo poco clara. Los tribunales apenas están comenzando a abordar las implicancias de las obras generadas por IA, dejando a los creadores vulnerables. Las leyes de derechos de autor actuales no están preparadas para manejar disputas de propiedad cuando las máquinas componen canciones utilizando fragmentos de miles de obras existentes.

Para nombrar el caso más reciente de la industria al momento de entregar este artículo, la sociedad alemana de gestión colectiva GEMA presentó una demanda contra Suno,[102] una empresa de inteligencia artificial valorada en 500 millones de dólares, alegando que la empresa procesa grabaciones protegidas sin autorización ni compensación. GEMA sostiene que la IA de Suno genera contenido musical similar a canciones originales de artistas como Alphaville, Kristina Bach, Lou Bega y Modern Talking, perjudicando a los creadores. Como prueba la actora realizó un análisis musicológico que demuestra las similitudes

101 Reed (2024).

102 GEMA, "Fair Remuneration Demanded: GEMA Files Lawsuit Against Suno Inc.", Comunicado de prensa, GEMA, 21 de enero de 2025, https://www.gema.de/en/w/press-release-lawsuit-against-suno.

entre el éxito *Daddy Cool* de Boney M. y la versión generada por IA. Este caso vuelve a destacar el problema central del uso de obras protegidas como material de entrenamiento para estos sistemas. Representantes de GEMA sostienen que la creatividad humana es la base de toda IA generativa, siendo su principal reclamo que el mercado de la IA ha carecido hasta ahora de principios fundamentales como transparencia, equidad y respeto.

En definitiva, el uso no autorizado de contenido protegido como insumo para la IA plantea una preocupación clave: la necesidad de establecer reglas y licencias que garanticen una cooperación justa entre los desarrolladores de IA y los titulares de derechos.

Otra preocupación que despierta la gran cantidad de contenido generado con IA que ha surgido en el mercado en el último tiempo tiene que ver con la transparencia. La transparencia es fundamental para no fomentar un sistema que explota la creatividad con fines de lucro. Si plataformas como Spotify y YouTube incorporan contenido generado por IA –quizás disfrazado como de creación humana–, se erosiona la confianza del oyente y se devalúa el arte genuino. Las plataformas de streaming pueden ahorrar miles de millones reemplazando música con licencias legítimas por alternativas generadas por IA. La transparencia es esencial para que no se desdibuje la línea entre el arte auténtico y la eficiencia corporativa,[103] por lo que es importante que las plataformas estén obligadas a informar cuando el contenido ha sido generado por IA.

El acceso a modelos de IA avanzados[104] seguirá expandiéndose, haciendo más fácil y accesible el desarrollo de herramientas de generación musical que rivalizan con los estudios profesionales.[105] Habiendo estado involucrados en la industria discográfica durante la llegada de Napster, fuimos testigos de cómo una tecnología disruptiva basada en la infracción masiva del derecho de autor puso en jaque a toda la industria. Sin embargo, también presenciamos cómo esta supo adaptarse y sobrevivir gracias a enfoques innovadores y la llegada de nuevas tecnologías.

103 Berger, V. AI's Impact On Music In 2025: Licensing, Creativity And Industry Survival, Forbes**,** 30 diciembre, 2024 https://www.forbes.com/sites/virginieberger/2024/12/30/ais-impact-on-music-in-2025-licensing-creativity-and-industry-survival/

104 Anderljung, Barnhart, Korinek, Leung, O'Keefe, & Whittlestone, et al, "Frontier AI Regulation: Managing Emerging Risks to Public Safety." 2023. https://arxiv.org/abs/2307.03718

105 Virginie Berger, "AI's Impact On Music In 2025: Licensing, Creativity And Industry Survival", *Forbes*, 30 de diciembre de 2024, https://www.forbes.com/sites/virginieberger/2024/12/30/ais-impact-on-music-in-2025-licensing-creativity-and-industry-survival/.

En el caso de Napster, el rescate de la industria discográfica vino con la irrupción del iPod en 2001 y la posibilidad de implementar un modelo de cobro por descarga, un cambio que permitió a las discográficas adaptarse al nuevo panorama. Posteriormente, con la irrupción del *streaming*, la industria ajustó su modelo de negocio para adaptarse. La clave en estos casos fue encontrar maneras efectivas de proteger los derechos de autor frente a tecnologías nuevas que, claramente, llegaron para quedarse. El desafío fue hallar modelos que permitieran a los titulares de derechos cobrar por el uso de sus obras en esos nuevos contextos.

El avance de la IA nos enfrenta a un desafío similar. Será esencial explorar soluciones que garanticen un equilibrio entre innovación tecnológica y protección de los derechos de autor. Por ejemplo, un primer paso interesante aparece en plataformas como *Fairly Trained*,[106] diseñadas para asegurarse de que los modelos de IA sean entrenados con contenido autorizado y respeten las licencias. A diferencia de lo que ocurrió en la época de Napster, hoy contamos con tecnologías como *blockchain* que puede ofrecer una oportunidad sin precedentes para registrar y rastrear el uso de las obras de un artista de manera transparente y confiable, las cuales permitirían por ejemplo implementar sistemas de retribución por el uso de obras protegidas en los *datasets* de entrenamiento. Con sistemas bien diseñados, los creadores podrían recibir compensaciones automáticas por el uso de sus obras, garantizando así un mecanismo eficiente para proteger y gestionar sus derechos en el entorno digital.

Habrá que observar cómo los titulares de derechos podrán aprovechar estas herramientas para defender y monetizar sus creaciones en un mundo tecnológico en constante evolución. Es evidente que la situación actual no solo restringe los derechos de las industrias creativas, sino que también puede comprometer la innovación y el crecimiento en el sector de la IA. Es imprescindible adoptar medidas adecuadas para fomentar el crecimiento, proteger la creatividad humana y generar confianza entre sectores.

En este trabajo hemos analizado las alternativas que los países de la región pueden considerar frente a los desafíos que surgen en relación con el uso de obras protegidas y datos en los sistemas de IA. Para ello, los cambios deben centrarse en tres aspectos clave: (i) garantizar que los creadores puedan controlar su contenido y recibir una compensación justa por su uso; (ii) impulsar el desarrollo de modelos de IA con acceso amplio y legal a datos de alta calidad; y (iii) establecer reglas claras para el uso de estas herramientas

106 Para más información ver https://www.fairlytrained.org/

protegiendo a su vez a los creativos de la nueva era. En esencia, el objetivo es encontrar una solución que beneficie tanto a las partes interesadas, como a la cultura y la economía.

IX. BIBLIOGRAFÍA

Anderljung, Barnhart, Korinek, Leung, O'Keefe, & Whittlestone, *et al*, "Frontier AI Regulation: Managing Emerging Risks to Public Safety." 2023. https://arxiv.org/abs/2307.03718

Annelise Levy, Aruni Soni, y Kyle Jahner, "OpenAI Bid to Centralize IP Suits Follows Discovery Setbacks", *Bloomberg Law*, 6 de diciembre de 2024, https://news.bloomberglaw.com/ip-law/openai-bid-to-centralize-ip-suits-follows-discovery-setbacks

Audrey Pope, "NYT V. OpenAI: The Times's About-Face", Harvard Law Review, 10 de abril de 2024, https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/

Aurea Suñol, "Aprovechamiento Indebido de la Reputación de una Marca y Free-riding", Almacén de Derecho, 1 de septiembre de 2015, https://almacendederecho.org/aprovechamiento-indebido-de-la-reputacion-de-una-marca-y-free-riding

Benjamin Hardman y James Housel, "A Sui Generis Approach To The Protection Of AI-Generated Works: Balancing Innovation And Authorship", *SSRN Electronic Journal*, 1 de enero de 2023, https://doi.org/10.2139/ssrn.4557004

Bernard Marr, "Cómo se puede explicar en forma sencilla la diferencia entre la Inteligencia artificial generativa y la tradicional", Forbes Argentina, 24 de julio de 2023, https://www.forbesargentina.com/innovacion/como-puede-explicar-forma-sencilla-diferencia-inteligencia-artificial-generativa-tradicional-n37661

Camilo Mirosevic Verdugo, "Origen y Evolución del Derecho de Autor Con Especial Referencia Al Derecho Chileno", *Revista de Derecho de La Pontificia Universidad Católica de Valparaíso* XXVIII (2007): 35-82, https://www.redalyc.org/articulo.oa?id=173620169001

Catherine Jewell, "En defensa del derecho de autor: visión de los interesados", WIPO - World Intellectual Property Organization, abril de 2014, https://www.wipo.int/wipo_magazine/es/2014/02/article_0004.html

César Huerta Ortiz, "Inteligencia Artificial Pone a Prueba a Filmes Mexicanos," El Universal, October 13, 2023, https://www.eluniversal.com.mx/espectaculos/ia-pone-a-prueba-a-filmes-mexicanos/

Charlotte Tobitt, "Who's suing AI and who's signing: Publisher deals vs lawsuits with generative AI companies", Press Gazette, 6 de diciembre de 2024, https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/

Christian Schmitz Vaccaro, "Evolucion de la Regulacion Internacional de la Propiedad Intelectual," Revista la Propiedad Inmaterial 17 (2013): 63-92

Chris Cooke, «Anthropic's Latest Legal Filing Tells Court To Focus On Whether Using Lyrics Is "Transformative Fair Use"», Complete Music Update, 19 de agosto de 2024, https://completemusicupdate.com/anthropics-latest-legal-filing-tells-court-to-focus-on-whether-using-lyrics-is-transformative-fair-use/

Delia Lipszyc, *Derecho de Autor y Derechos Conexos* (UNESCO-CERLALC-ZAVALÍA, 1993)

Eduardo De La Parra Trujillo, "Nociones básicas sobre el objeto de los derechos de autor", *Revista Del Instituto de La Judicatura Federal* 18 (2004), https://escuelajudicial.cjf.gob.mx/publicaciones/revista/18/r18_4.pdf

EFE, "Los Principales Medios de Canadá Demandan A OpenAI Por Violar Sus Derechos de Autor", Telemundo Chicago, 29 de noviembre de 2024, https://www.telemundochicago.com/noticias/tecnologia/medios-canada-demandan-openai-violar-derechos-autor/2510377/

Elliot Samuel Paul y Dustin Stokes, "Creativity", Stanford Encyclopedia Of Philosophy, 16 de febrero de 2023, https://plato.stanford.edu/entries/creativity/#Spon

Greg Bensinger, "Focus: ChatGPT launches boom in AI-written e-books on Amazon", Reuters, 21 de febrero de 2023, https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/

Henry Shevlin, "Non-human Consciousness And The Specificity Problem: A Modest Theoretical Proposal", *Mind & Language* 36, n.o 2 (18 de enero de 2021): 297-314, https://doi.org/10.1111/mila.12338

Isabela Raygoza, "Bad Bunny Critica Canciones Que Lo Imitan Con IA Que Se Hacen Virales En TikTok", Billboard, 11 de septiembre de 2023, https://www.billboard.com/espanol/musica/bad-bunny-critica-canciones-generadas-ia-inteligencia-artificial-1235467073/

Joe Coscarelli, "An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World", The New York Times, 19 de abril de 2023, https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html

Leire Gutierrez Vázquez, "República Checa: Un Tribunal Se Pronuncia Sobre la Protección Por el Derecho de Autor de los Resultados De Inteligencia Artificial", Instituto Autor, 15 de mayo de 2024, https://institutoautor.org/republica-checa-un-tribunal-se-pronuncia-sobre-la-proteccion-por-el-derecho-de-autor-de-los-resultados-de-inteligencia-artificial/

Lessig, L, Cultura libre: Cómo los grandes medios usan la tecnología y las leyes para encerrar la cultura y controlar la creatividad, LOM Ediciones, Primera edición en Chile, 2005. I.S.B.N: 956-282-745-3. p.148 (Título Original: "Free Culture: How Big Media uses Technology and the Law to lock down Culture and control Creativity")

Margaret A. Boden, "Creativity And ALife", Artificial Life 21, n.o 3 (1 de agosto de 2015): 354-65, https://doi.org/10.1162/artl_a_00176

Mark A. Lemley y Bryan Casey, "Fair Learning", Texas Law Review 99, n.o 4 (30 de enero de 2020): 743-85, https://doi.org/10.2139/ssrn.3528447

Micaela Mantegna, "ARTificial: Why Copyright Is Not the Right Policy Tool to Deal With Generative AI," The Yale Law Journal 133 (April 22, 2024): 1126, https://www.yalelawjournal.org/forum/artificial-why-copyright-is-not-the-right-policy-tool-to-deal-with-generative-ai

Muhammad Abbas, Farooq Ahmed Jam, y Tariq Iqbal Khan, "Is It Harmful Or Helpful? Examining The Causes And Consequences Of Generative AI Usage Among University Students", International Journal Of Educational Technology In Higher Education 21 (16 de febrero de 2024), https://doi.org/10.1186/s41239-024-00444-7

Nantheera Anantrasirichai y David Bull, "Artificial Intelligence In The Creative Industries: A Review", Artificial Intelligence Review 55, n.o 1 (2 de julio de 2021): 589-656, https://doi.org/10.1007/s10462-021-10039-7

Noemí Pulido Pavón, Luis Palma Martos, y Luis F. Aguado, "Derechos de Autor. Enfoque Económico, Evolución y Perspectivas", *Revista de Economía Institucional* 18, n.o 35 (29 de noviembre de 2016): 151, https://doi.org/10.18601/01245996.v18n35.08

Patricia Esteban, "Un Tribunal Alemán, Por Primera Vez En Europa, Rechaza el Recurso de un Fotógrafo Al Que una Inteligencia Artificial Le Robó Su Foto", Cinco Días, 3 de octubre de 2024, https://cincodias.elpais.com/legal/2024-10-03/un-tribunal-aleman-por-primera-vez-en-europa-rechaza-el-recurso-de-un-fotografo-al-que-una-inteligencia-artificial-le-robo-su-foto.html

Rachel Reed, "ChatNYT", Harvard Law Today, 22 de marzo de 2024, https://hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/

Rubén Darío Echeverri Echeverri y Luz Marina Franco Montoya, "La propiedad intelectual en el marco de los derechos de propiedad", *Teuken Bidikay: Revista Latinoamericana de Investigación En Organizaciones, Ambiente y Sociedad*, ISSN-e 2619-1822, ISSN 2215-8405, Vol. 1, N° 1 (2010): págs. 33-51, https://dialnet.unirioja.es/servlet/articulo?codigo=8760977

Sundara Rajanm, Mira T. Is Generative AI Fair Use of Copyright Works? NYT v. OpenAI,Kluwer Copyright Blog, Consultado 28 Noviembre, 2024, https://copyrightblog.kluweriplaw.com/2024/02/29/is-generative-ai-fair-use-of-copyright-works-nyt-v-openai/#:~:text=Crucially%2C%20the%20use%20of%20NYT,the%20doctrine%20of%20fair%20use

Tony Berber Sardinha, "AI-generated Vs Human-authored Texts: A Multidimensional Comparison", *Applied Corpus Linguistics* 4, n.o 1 (20 de diciembre de 2023): 100083, https://doi.org/10.1016/j.acorp.2023.100083

Virginie Berger, "AI's Impact On Music In 2025: Licensing, Creativity And Industry Survival", *Forbes*, 30 de diciembre de 2024, https://www.forbes.com/sites/virginieberger/2024/12/30/ais-impact-on-music-in-2025-licensing-creativity-and-industry-survival/

Yogesh K. Dwivedi et al., «Opinion Paper: "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives On Opportunities, Challenges And Implications Of Generative Conversational AI For Research, Practice And Policy», International Journal Of Information Management 71 (agosto de 2023): 102642, https://doi.org/10.1016/j.ijinfomgt.2023.102642

GENERAL FRAMEWORK OF REGULATION IN THE PEOPLE'S REPUBLIC OF CHINA ON ARTIFICIAL INTELLIGENCE AND DIGITAL TRANSFORMATION

SUMMARY

I. Introduction **II.** Laws regulating AI in China **III.** Measures regulating AI in China **IV.** Provisions on the management of algorithmic recommendations in internet information services **V.** Other sectoral AI regulations and technical standards in China **VI.** AI legislation proposals by leading experts **VII.** Conclusion **VIII.** Bibliography

> HUI ZHOU

Deputy director, associate professor. Department of cyber and information law, Institute of law, Chinese Academy of Social Sciences

> YUE ZHU

Assistant professor of law at Tongji University

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has significantly transformed various sectors worldwide, prompting governments to develop regulatory frameworks that address both the opportunities and challenges presented by this technology. In China, the government has been proactive in establishing various laws and regulations to guide the development, deployment, and use of AI. These legal instruments aim to promote innovation while safeguarding national security, protecting personal rights, and ensuring ethical standards are upheld.

This brief provides an in-depth examination of China's AI regulatory landscape, beginning with foundational laws such as the CSL, DSL, and PIPL. These laws set the groundwork for data handling, personal information protection, and cybersecurity, which are crucial for the responsible development of AI technologies.

Further, the brief explores specific measures tailored to address the nuances of AI technologies, including the Provisions on the Management of Algorithmic Recommendations in Internet Information Services, the Provisions on the Administration of Deep Synthesis Internet Information Services, the Interim Measures for the Management of Generative Artificial Intelligence Services, and the Measures for the Labeling of AI-Generated Synthetic Content (official draft for suggestion). These regulations focus on algorithmic transparency, content authenticity, and the ethical deployment of AI applications.

Additionally, the brief quite sketchily presents sector-specific regulations and technical standards that guide AI applications in industries such as finance, healthcare, transportation, and education. These sectoral regulations complement the broader legal framework by addressing industry-specific risks and requirements, ensuring that AI technologies are applied responsibly within different contexts.

Finally, the brief discusses proposed AI legislation, namely the MAIL and AIL, which offer insights into the potential future direction of AI governance in China. These proposals similarly emphasize ethical considerations, innovation promotion, and the protection of individual rights, reflecting China's consensual commitment to shaping a legal environment conducive to responsible AI development.

By analyzing all these instruments, the brief aims to provide a comprehensive understanding of how China navigates the complex landscape of AI governance. It highlights the country's efforts to balance the encouragement of technological innovation with the imperative to protect various rights and interests of state, society, and individuals, thereby contributing to the global discourse on AI regulation.

II. LAWS REGULATING AI IN CHINA

In China, several laws have been enacted to regulate various aspects of AI, particularly focusing on network and data security, and personal information protection. This section examines the AI-related provisions within three key Chinese laws: As mentioned above, the CSL, the DSL, and the PIPL. These laws collectively shape the regulatory landscape for AI in China by establishing guidelines for data handling, personal information protection, and the legitimate application of AI.

1. CSL

The Cybersecurity Law, effective since June 2017, serves as the foundational legislation governing cybersecurity in China. While it does not explicitly mention AI, its provisions have significant implications for AI systems, particularly concerning data security, network security, and the obligations of network operators.

Key AI-Relevant Provisions:

Obligations of Network Operators: Network operators must implement security measures to protect their networks and data from unauthorized access, interference, or destruction. They are required to adopt technical measures to prevent computer viruses and cyber-attacks. For AI systems, this means that certain entities developing and/or deploying AI technologies must ensure robust cybersecurity measures are in place to protect data integrity and prevent malicious exploitation of AI algorithms.

Network Product Security: Providers of network products and services shall ensure compliance with mandatory requirements set forth by national standards and must not include malicious programs. When security flaws or vulnerabilities are discovered, they must immediately adopt remedial measures and inform users. AI products and services covered herein shall adhere to these security standards, ensuring the integrity and reliability of AI applications.

Prohibition of Malicious Activities: Individuals and organizations are prohibited from engaging in certain activities that harm cybersecurity, including providing programs or tools used for network intrusions. This may encompass AI tools, including frontier AI tools, that could be used maliciously, potentially placing legal responsibilities on developers and distributors of AI technologies.

Critical Information Infrastructure Security: Operators of critical information infrastructure (CII) have additional security obligations, including conducting security assessments and protecting important data. AI systems classified as part of CII must implement enhanced security measures, potentially affecting sectors like finance, transportation, and healthcare where AI is increasingly integrated.

Implications for AI:

The CSL establishes a legal framework that affects AI development and deployment in China. AI systems shall therefore comply with stringent cybersecurity measures. AI companies must ensure that their network and data handling practices align with the obligations set forth in the law. The emphasis on network and data security directly impacts how AI technologies are developed, implemented, and managed.

2.DSL

Effective from September 2021, the Data Security Law provides a comprehensive framework for data handling and security in China. While not exclusively focused on AI, the law's emphasis on data governance directly impacts AI systems that rely heavily on data collection and processing.

Key AI-Relevant Provisions:

Definition of Data Handling: The law defines data handling as the collection, storage, use, processing, transmission, provision, and disclosure of data. This broad definition may encompass many AI activities involving data, making AI systems subject to the law's requirements.

Categorized and Graded Data Protection System: The law establishes a system to categorize and grade data based on its importance to national security, public interests, or individual rights. AI systems processing important or core national data must adhere to stricter security measures, including enhanced protection protocols and potential restrictions on data handling practices.

Promotion of Data Development and Use: The law encourages the advancement of data infrastructure and supports innovative applications of data in all industries. It promotes research into data development and use technologies, fostering AI innovation while ensuring data security. AI companies can potentially leverage state support for research and development, aligning their innovation strategies with national priorities.

Data Security Obligations: Entities handling data must establish data security management systems and adopt necessary technical measures to ensure data security. AI companies must implement robust data protection mechanisms throughout the data lifecycle, including data encryption, access controls, and regular security assessments.

The law also establishes systematic rules for data security risk assessment, reporting, information sharing, monitoring, and early warning.

Implications for AI:
The Data Security Law imposes requirements on data handling practices, directly affecting AI systems that depend on large datasets. AI operators covered shall accordingly implement comprehensive data security measures and comply with data classification and grading requirements. The law seeks to balance the promotion of data-driven innovation with the protection of national security and individual rights. AI companies must be vigilant in their data governance strategies to align with these legal obligations.

3. PIPL

In effect since November 2021, the PIPL is China's first comprehensive law dedicated to personal data protection. It introduces specific provisions related to AI, particularly concerning automated decision-making and the rights of individuals.

Key AI-Relevant Provisions:
Definition of Personal Information and Automated Decision-Making: Personal information is defined as information related to identified or identifiable natural

persons. Automated decision-making is defined as the use of \programs to automatically analyze or assess personal behaviors and make decisions based on that analysis. AI systems that process personal data for decision-making probably fall under these definitions and are subject to the law's provisions.

Transparency and Fairness in Automated Decision-Making: Personal information handlers using automated decision-making must ensure transparency, fairness, and justice in their processes. They must not engage in unreasonable differential treatment of individuals in terms of trade conditions such as pricing. This provision addresses concerns about AI algorithms potentially leading to biased or unfair outcomes, mandating that AI systems be designed and operated in a manner that is equitable.

Right to Explanation and Refusal: When automated decision-making produces decisions with a significant impact on individuals' rights and interests, individuals have the right to request an explanation and to refuse the processing based solely on automated decision-making. This empowers individuals to challenge AI-driven decisions that affect them, compelling AI developers to provide mechanisms for human oversight and intervention.

Consent for Sensitive Personal Information: Handling sensitive personal information generally requires obtaining separate consent from individuals. AI systems processing such data must ensure they have the necessary consent and implement strict protection measures, including enhanced security protocols and access restrictions.

Obligations for Personal Information Handlers: Entities handling personal information must adopt measures to prevent unauthorized access, leaks, or misuse. They must conduct personal information protection impact assessments, particularly when using personal data for automated decision-making. Large platform operators with complex business models have additional obligations, such as establishing independent bodies to supervise personal information protection and publishing social responsibility reports. This promotes greater accountability and transparency in AI operations.

Cross-Border Data Transfer: The PIPL imposes conditions on transferring personal information outside of China, including passing security assessments or obtaining certifications. AI systems that rely on cross-border data processing must comply with these requirements, potentially impacting international AI collaborations and data exchange.

Implications for AI:
The PIPL has significant implications for AI systems that handle personal data. It emphasizes individuals' rights over their personal information, including the right to explanation and refusal of automated decision-making. AI developers in many circumstances must ensure transparency in their algorithms, avoid biases, and implement mechanisms to allow individuals to exercise their rights. The stringent consent requirements for sensitive data necessitate careful data handling practices. Furthermore, AI companies must establish robust compliance programs, conduct regular audits, and stay informed of legal developments to mitigate risks.

Combined Implications for AI:
The interplay of these three laws creates a comprehensive regulatory environment for AI in China. AI systems must navigate complex requirements. Key considerations may include:

Data governance. AI relies on vast amounts of data. Companies must implement robust data management systems, ensure data quality, and comply with data classification and protection requirements. This includes adhering to the categorized and graded protection system for data and ensuring secure data storage and processing.

Algorithmic transparency and fairness. The emphasis on transparency and fairness in automated decision-making requires AI developers to address algorithmic biases and provide explanations for AI-driven decisions, aligning with global concerns about ethical AI. AI systems must be designed to prevent discrimination and allow individuals to understand and challenge decisions affecting them.

Cross-border data challenges. Restrictions on cross-border data transfer impact AI collaborations and operations involving international data flows. Companies must assess their data transfer practices and explore mechanisms such as security assessments or obtaining certifications to legally transfer data abroad if necessary for global collaboration.

Innovation, While the laws impose obligations, they also support AI innovation through state initiatives. Companies can align their research and development efforts with national strategies to leverage support and contribute to China's AI development goals. The promotion of data development and use technologies provides opportunities for AI advancement within the legal framework.

III. MEASURES REGULATING AI IN CHINA

In addition to the foundational laws governing cybersecurity, data security, and personal information protection in China, several specific measures have been introduced to further regulate the development and deployment of AI. These measures address AI in specific contexts such as algorithmic recommendation, deep synthesis, generative AI, and the labeling of AI-generated content. This section provides a comprehensive summary of four key regulatory instruments: the Provisions on the Management of Algorithmic Recommendations in Internet Information Services, the Provisions on the Administration of Deep Synthesis Internet Information Services, the Interim Measures for the Management of Generative Artificial Intelligence Services, and the Measures for Labeling of AI-Generated Synthetic Content (official draft for suggestion). These measures collectively aim to promote the healthy and orderly development of AI in China while safeguarding national security, social order, and individual rights.

IV. PROVISIONS ON THE MANAGEMENT OF ALGORITHMIC RECOMMENDATIONS IN INTERNET INFORMATION SERVICES

Effective from March 1, 2022, the Provisions on the Management of Algorithmic Recommendations in Internet Information Services establish a regulatory framework for the use of algorithmic recommendation technologies in internet information services. These technologies may include algorithms that generate and synthesize content, provide individualized pushes, filter searches, and make scheduling decisions to offer users information tailored to their preferences.

Key AI-Relevant Provisions:

Compliance and Ethical Use: Service providers must adhere to laws and regulations, respect social norms and ethics, and follow principles of fairness, transparency, and accountability. They are required to prevent the use of algorithmic recommendations for activities that endanger national security, disrupt social order, or infringe on the lawful rights of others.

Entity Responsibility and Management Systems: Providers must implement entity responsibility for algorithm security, establishing management systems for algorithm mechanism checks, scientific and technological ethics reviews, user registration, data security, and personal information protection. They must

allocate professional staff and technical support proportional to the scale of their services.

Algorithm Review and Assessment: Providers are required to periodically check, assess, and verify algorithm mechanisms, models, data, and outcomes. They must ensure that algorithms do not violate laws or ethical standards, such as inducing users to become addicted or overspend.

Information Security Management: Providers must strengthen the management of information security by creating a database of characteristics used to identify illegal and negative content. They must prevent the transmission of prohibited information and ensure that algorithmically generated or synthesized content is conspicuously labeled if required.

User Modeling and Tag Management: Providers must manage user models and tags responsibly, ensuring that unlawful or negative information is not used to profile users or push content to them.

Transparency and User Choice: Providers must inform users conspicuously about the algorithmic recommendation services they offer, including basic principles and main operation mechanisms. They must provide users with options to disable personalized recommendations and to select or delete user tags that influence the algorithmic recommendations.

Protection of Special Groups: Special considerations are mandated for services provided to minors, seniors, and laborers. For minors, providers must prevent content that could negatively impact their physical and psychological health and avoid inducing addiction. For seniors, services must consider their needs and safeguard their rights, including protection against fraud. For laborers, providers must ensure that algorithms related to work assignments and compensation protect lawful rights such as fair wages and reasonable working hours.

Fair Transaction Rights: In commercial transactions, providers must not use algorithms to conduct unreasonable differential treatment of consumers, such as discriminatory pricing based on consumer preferences or behavior.

Oversight and Filing Requirements: Certain poviders must perform filing procedures, including submitting details about their algorithmic services. Authorities establish a management system for oversight.

Implications for AI:

This measure places significant responsibility on AI developers and service providers to ensure that their algorithmic recommendation systems are transparent, ethical, and compliant with legal standards. It emphasizes user rights,

including the right to understand and control how algorithms influence the information they receive. AI systems must also avoid discriminatory practices and protect vulnerable groups, aligning with broader concerns about algorithmic fairness and accountability.

Provisions on the Administration of Deep Synthesis Internet Information Services

Effective from January 10, 2023, the Provisions on the Administration of Deep Synthesis Internet Information Services specifically regulate technologies known as "deep synthesis," which include deep learning and virtual reality techniques used to generate or manipulate content such as text, images, audio, and video. This includes technologies that can create realistic virtual scenes or manipulate or alter biometric information, often referred to in the context of "deepfakes."

Key AI-Relevant Provisions:

Prohibition of Illegal Content and Activities: Service providers and users are prohibited from using deep synthesis technologies to produce or disseminate content that is illegal, harmful to national security, or infringes upon the rights of others. Specifically, they must not create or spread fake news using deep synthesis technologies.

Information Security Responsibility: Providers must implement primary responsibility for information security, establishing management systems for user registration, algorithm review, data security, and personal information protection. They must have safe and controllable technical safeguard measures in place.

Real Identity Verification: Providers must verify the real identity of users through methods such as mobile numbers or identification documents, ensuring accountability and traceability of users who access deep synthesis services.

Content Review and Management: Providers must strengthen the management of content generated through deep synthesis, reviewing user inputs and synthesis outcomes. They must establish a database of characteristics to identify illegal and negative content and take appropriate measures when such content is discovered.

Mechanisms for Dispelling Rumors: Providers must establish mechanisms to address and dispel misinformation or rumors generated through deep synthesis technologies, reporting such incidents to relevant authorities.

Training Data Management: Providers and technical supporters must manage training data responsibly, ensuring data security and compliance with

personal information protection laws. If biometric information is involved, they must obtain consent from the individuals concerned.

Security Assessments: Providers offering tools with certain functionalities, such as generating or editing biometric information, must conduct security assessments either independently or through professional bodies.

Mandatory Labeling: Providers must employ technical measures to attach identifiers to content generated or edited using deep synthesis technologies. Prominent labels are required for certain types of content that might cause confusion or mislead the public, such as AI-generated text, voice, or images that simulate real people.

User Appeals and Complaints: Providers must set up accessible channels for user appeals, complaints, and reports, and must handle them promptly, providing feedback on the outcomes.

Implications for AI:

The provisions address the potential misuse of AI technologies that can create realistic synthetic content, especially so-called "deepfake" content, which poses risks such as misinformation, fraud, and infringement of personal rights. AI developers and service providers must ensure transparency through labeling, obtain necessary consents, and implement robust security and ethical measures in their technologies. The measures aim to balance innovation with the need to prevent the negative societal impacts of deep synthesis technologies.

Interim Measures for the Management of Generative Artificial Intelligence Services

Effective from August, 2023, the Interim Measures for the Management of Generative Artificial Intelligence Services provide regulatory guidelines for the development and use of generative AI technologies that produce content such as text, images, audio, and video for public consumption.

Key AI-Relevant Provisions:

Scope and Application: The measures apply to the use of generative AI technologies to provide services to the public within China. They emphasize the balance between promoting innovation and ensuring security, indicating a tolerant and cautious approach to regulation.

Compliance and Ethical Standards: Providers must adhere to laws and administrative regulations, respect social ethics, and uphold mainstream social values. They must prevent the generation of prohibited content, such

as that which endangers national security, promotes extremism, or contains misinformation.

Non-Discrimination: During all stages of AI development and deployment, providers must prevent discrimination based on factors like race, ethnicity, gender, or health. This aligns with global concerns about AI fairness and the prevention of biased outcomes.

Intellectual Property and Fair Competition: Providers must respect intellectual property rights and commercial ethics, avoiding monopolistic practices or unfair competition by exploiting advantages in algorithms or data.

Protection of Personal Rights: Providers must prevent the infringement of others' lawful rights, including rights to image, reputation, privacy, and personal information. They must avoid content that could harm the physical or psychological well-being of individuals.

Data and Model Requirements: Providers must ensure that the data and foundational models used in training have lawful sources. They must respect intellectual property rights and obtain consent when personal information is involved. Measures must be taken to improve the quality, truthfulness, and diversity of training data.

User Privacy and Data Protection: Providers are responsible for the confidentiality of user input and usage records. They must not collect unnecessary personal information or unlawfully provide user data to others. They are obligated to respond to user requests regarding personal information access, correction, or deletion.

Content Labeling: Providers must label AI-generated content, especially images and videos, in accordance with existing provisions on deep synthesis services. This ensures transparency and allows users to distinguish between AI-generated and human-generated content.

Complaint and Reporting Mechanisms: Providers must establish mechanisms for handling public complaints and reports, ensuring that issues are addressed promptly and feedback is provided.

Security Assessments and Filings: Certain providers must conduct security assessments and comply with filing requirements.

Implications for AI:

These measures place comprehensive obligations on providers of generative AI services to ensure legal compliance, ethical conduct, and the protection of individual rights. For example, the emphasis on data quality and lawful sourcing addresses concerns about AI models producing inaccurate or harmful content

due to poor training data. Providers are encouraged to innovate responsibly while being held accountable for the societal impacts of their technologies.

Measures for Labeling of AI-Generated Synthetic Content (Official Draft for Suggestion)

As a draft measure not yet in effect, the Measures for Labeling of AI-Generated Synthetic Content aim to standardize the labeling practices for AI-generated content, enhancing transparency and protecting users from potential deception or misinformation.

Key AI-Relevant Provisions:

Scope and Definitions: The measures apply to online information service providers offering services that generate synthetic content using AI technologies. AI-generated synthetic content includes text, images, audio, video, or other information produced using AI.

Labeling Requirements: Service providers must add both explicit and implicit labels to AI-generated content. Explicit labels are perceivable by users and must be added in appropriate positions depending on the content type (e.g., text notifications for text content, conspicuous symbols on images). Implicit labels are embedded in the metadata of content files, including information such as the provider's name and content reference number.

Verification and Regulation by Platforms: Online content transmission platforms must verify whether uploaded content contains implicit labels and add conspicuous notification labels when necessary. They are responsible for alerting users if content is AI-generated or suspected to be so.

User Obligations: Service providers must explain labeling methods and specifications in user agreements, ensuring users are aware of their obligations. Users are prohibited from maliciously deleting, altering, or concealing labels and must not harm others' rights through improper labeling.

Compliance with Standards: Service providers must conduct labeling in accordance with another mandatory national standards, which is also soliciting public suggestion, therefore ensuring consistency and compliance across the industry.

Implications for AI:

By mandating clear labeling of AI-generated content, the draft measures aim to enhance transparency and allow users to identify and assess content appropriately. This is particularly important in combating misinformation and

rumor and ensuring that users are not misled by synthetic content that appears authentic. The measures encourage responsible use of AI technologies and hold both service providers and users accountable for maintaining transparency.

Combined Implications for AI

The four measures collectively reinforce China's commitment to regulating AI technologies in a manner that promotes innovation while safeguarding national security, social order, and individual rights. Key themes across these measures include:

Transparency and Accountability. Providers of AI services are required to be transparent about their algorithms, data sources, and the nature of the content they generate or recommend. Mandatory labeling of AI-generated content ensures that users are aware when they are interacting with synthetic media.

Content Management and Security: Providers are responsible for preventing the dissemination of illegal or harmful content, including misinformation, extremism, and content that could negatively impact vulnerable groups such as minors and seniors. Robust content review mechanisms and security measures are mandated.

Ethical Use of AI Technologies: The measures stress adherence to ethical standards, including respect for intellectual property rights, prevention of unfair competition, and avoidance of practices that could lead to addiction or exploitation of users.

Protection of Individual Rights. There is a strong emphasis on protecting personal information, preventing discrimination, and safeguarding the lawful rights and interests of individuals. AI systems must be designed and operated to respect privacy, image rights, and prevent harm.

Regulatory Oversight and Compliance: Providers must comply with filing requirements, security assessments, and cooperate with regulatory authorities. Non-compliance can result in penalties, including fines and suspension of services.

V. OTHER SECTORAL AI REGULATIONS AND TECHNICAL STANDARDS IN CHINA

In addition to the overarching laws and specific measures previously discussed, China has implemented, and is enacting, various sector-specific legal instruments and technical standards to further guide the development and application of AI across different industries. These sectoral regulations complement the broader legal framework established by the laws and measures mentioned above. A surely incomplete list of the most important contexts include:

1. Financial Sector Regulations. Banks, insurance companies, and other financial institutions are encouraged to use AI to improve service efficiency while ensuring compliance with cybersecurity and data protection laws. Strict security measures are required when using AI technologies for activities like credit scoring, fraud detection, and personalized financial services. Compliance with data localization and cross-border data transfer regulations is crucial.
2. Healthcare Sector Regulations. AI systems used for medical and diagnostic support must meet stringent safety standards, and patient data must be protected in accordance with personal information protection laws. It is worthwhile to mention that the National Medical Products Administration provides guidance on the registration and evaluation of AI-based medical devices, ensuring they meet quality and safety requirements before entering the market.
3. Autonomous Vehicles and Transportation. The Ministry of Industry and Information Technology has issued policies to promote the development of autonomous driving technologies. Regulations focus on safety testing, data security, and the protection of passenger information. Companies testing autonomous vehicles must obtain necessary permits and ensure compliance with traffic laws. Besides, several local governments, such as Beijing, have implemented specific guidelines for the testing and deployment of autonomous vehicles on public roads, including requirements for safety drivers and reporting of test results.

Technical standards also play a crucial role in guiding the development and deployment of AI technologies. China has been actively developing national standards to ensure consistency, safety, and interoperability of AI systems. China also actively participates in international standardization organizations like the International Organization for Standardization

(ISO) and the International Electrotechnical Commission (IEC), contributing to global AI standards and aligning domestic standards with international practices.

All these sectoral regulations and technical standards complement the general laws and measures previously discussed. They provide detailed requirements tailored to specific applications of AI, ensuring that industry-specific risks are managed effectively. Together, they form a comprehensive, well-balanced, and indeed workable regulatory ecosystem. Entities operating within these sectors must not only comply with the general laws and measures but also adhere to the specific regulations and technical standards relevant to their industry. This multi-layered approach ensures that AI technologies are developed and applied in a manner that is consistent with China's policy objectives and legal requirements. As AI continues to evolve, these regulations and standards will play a crucial role in guiding stakeholders toward sustainable and socially beneficial AI innovation.

VI. AI LEGISLATION PROPOSALS BY LEADING EXPERTS

Chinese legal experts have proposed comprehensive legislative frameworks to guide the responsible development and application of AI. Two notable expert proposals are the MAIR and AIR. While neither is a binding legal instrument, they offer valuable insights into the potential future direction of AI regulation in China. This section provides a summary of these two proposals, and a comparative analysis of their key features.

Overview of the MAIL

The MAIL is the earliest expert draft proposal aiming to suggest for a comprehensive legal framework for AI governance in China.

Key Provisions:

Scope and Application: The law applies to AI-related activities within China and extends to activities outside China that affect its national security or the rights of its citizens.

Principles of AI Governance: It outlines several guiding principles, including: Human-Centric Approach, AI should promote human welfare and remain under human supervision and control; Safety and Security, necessary measures must be taken to ensure the safety and security of AI systems and related data; Openness and Transparency, AI providers should appropriately label AI systems

and clarify their purposes, principles, and effects; Accountability, AI developers, providers, and users are responsible for their respective activities; Fairness and Equality, AI should avoid unreasonable differential treatment and consider the needs of special groups like minors and the elderly; and, Green Development, encouraging the use of energy-saving technologies in AI development.

Promotion of AI Development: Infrastructure Support, the state supports building AI infrastructure, including public computing resources and data centers; Algorithm and Model Innovation, encouraging open-source platforms and the protection of foundational models; Data Supply, supporting constructing foundational and specialized databases, promoting data sharing and ensuring data quality; Industrial Development, accelerating the application of key AI technologies in various industries; and, Talent Cultivation, supporting educational institutions in improving AI-related disciplines and research.

Regulatory Framework: Negative List Oversight System, establishing an "AI negative list" for AI applications that require licensing due to potential risks; Licensing Requirements, outlining conditions for obtaining licenses, including legal entity status, security measures, and compliance systems; Obligations of AI Developers and Providers, specifying safety obligations, risk management, transparency, explainability, fairness, and ethical review requirements; and, Special Obligations for Foundation Models, developers of foundational AI models have additional responsibilities, such as ensuring security, openness, and assisting others in compliance.

Comprehensive Governance Mechanism: Regulatory Authorities, establishing a China Administration of Artificial Intelligence as the main authority for AI governance; Collaborative Governance, promoting a governance mechanism involving government, corporations, industry self-regulation, and user self-discipline; and, Expert Committees, establishing national and local committees of experts on AI ethics.

Liability and Penalties: Administrative Penalties, specifying fines and corrective measures for violations; Civil Liability, outlining tort liability for damages caused by AI systems; Exemptions, providing liability exemptions for open-source AI providers under certain conditions; and, International Cooperation, encouraging participation in international AI rule-making and cooperation, aiming to develop global governance frameworks.

Overview of the AIL

The AIL is another influential proposal by Chinese academics that seeks to establish a legislative framework for AI governance.

Key Provisions:

Scope and Application: Applies to AI development, provision, and use within China, and to activities outside China that affect its national security or citizens' rights.

AI Regulation Principles: Scientific and Technological Ethics, AI development should be people-centered, respect personal freedom and dignity, and promote public welfare; Innovative Development, implementing inclusive and prudent supervision, encouraging AI industry innovation while ensuring safety; Fairness and Impartiality, Protecting legitimate rights and interests, avoiding unreasonable differential treatment; Transparency and Explainability, requiring AI to be transparent and explainable, providing users with necessary information; Safety and Accountability, AI entities must adopt measures to ensure safety and bear corresponding legal liabilities; and, Proper Use, AI must comply with laws, respect ethics, and not harm national security or public interests.

Promotion of AI Development: Infrastructure Development, strengthening AI infrastructure, including computing resources and public data platforms; Intellectual Property Protection, protecting AI-related intellectual property rights and encourages innovation; Open Source Ecosystem, supporting building open-source platforms and communities; Data Utilization, encouraging data sharing and reasonable use of data for AI development; and, Industry-Academia Integration, promoting collaboration between enterprises, educational institutions, and research bodies.

Protection of Users' Rights and Interests: Equal Rights, ensuring universality and fairness in AI products and services; Right to Know, providing users with clear information about AI services, including functions, purposes, and risks; Privacy Protection, safeguarding personal information and prohibits unlawful data retention and provision to others; Right to Explanation and Refusal, users can request explanations of AI decisions and have the right to refuse or withdraw from AI services; and, Protection of Vulnerable Groups, considering the needs of minors, the elderly, and people with disabilities.

Regulatory Framework: Main Oversight Departments, assigning responsibilities to relevant government departments for AI supervision; Grading and Categorization, establishing a system to grade and categorize AI based on importance and potential harm; Special Supervision of Critical AI, implementing special management for critical AI applications, such as those affecting personal rights or used in critical infrastructure; Risk Monitoring and Emergency Response, developing mechanisms for AI security monitoring and incident response.

Obligations of AI Entities: Security Obligations, AI developers and providers must ensure safety, conduct risk assessments, and report major incidents; Content Security, must prevent the generation of illegal or harmful content and add identifiers to AI-generated content. Compliance Guidelines: Ensure certain products and services comply with laws and regulations.

Special Application Contexts, including but not limited to: Use by State Organs, AI decisions can only be references for administrative actions; Judicial AI: AI assists in judicial work but cannot replace human judgment; Medical AI: AI decisions serve as references in medical activities; Autonomous Driving: Requires appropriate licenses and prioritizes safety.

Liability and Penalties: Administrative Penalties, specifying fines and corrective measures for violations; Civil Liability, outlining tort liability for damages, including punitive damages in severe cases; and Safe Harbor Rules, providing conditions under which AI service providers are exempt from liability.

Preliminary Comparative Analysis

While both proposals aim to establish comprehensive AI legislation in China, there are a few notable consensuses and differences in their approaches and emphases.

Similarities:

- Comprehensive Scope: Both laws cover AI development, provision, and use within China and extend to activities affecting China from abroad.
- Guiding Principles: Both emphasize ethical considerations, including human-centric development, safety, transparency, fairness, and accountability.
- Promotion of Innovation: Both support AI innovation, infrastructure development, data sharing, and open-source ecosystems.
- Protection of Rights: Both prioritize the protection of individual rights, including privacy, personal information, and the rights of vulnerable groups.
- Regulatory Oversight: Both establish regulatory frameworks involving government authorities responsible for AI governance.
- Liability Provisions: Both outline legal liabilities for AI developers, providers, and users, including administrative penalties and civil liabilities.
- International Cooperation: Both recognize the importance of participating in global AI governance and promoting international cooperation.

Differences:

Regulatory Structure:

- MAIL: Establishes a specific authority, the China Administration of Artificial Intelligence (CAAI), as the main regulatory body, and introduces a Negative List oversight system requiring licenses for certain AI applications.
- AIL: Distributes regulatory responsibilities among relevant government departments without specifying a central authority and emphasizes a grading and categorization system for AI oversight.

Licensing and Oversight Mechanisms:

- MAIL: Requires licenses for AI applications on the Negative List and details the licensing process, conditions, and obligations.
- AIL: Focuses on registration and special supervision for critical AI but does not specify a licensing system like the Negative List.

Obligations for Foundation Models:

- MAIL: Imposes specific obligations on developers of foundational AI models, including security measures, openness, and assisting others in compliance.
- AIL: Mentions foundation models in the context of critical AI but provides less detail on specific obligations.

Promotion of Open Source:

- MAIL: Encourages open-source development and the establishment of open-source AI foundations.
- AIL: Supports building open-source platforms but places more emphasis on intellectual property protection and reasonable use of data.

Special Application Scenarios:

- MAIL: Does not elaborate extensively on specific application scenarios but includes general provisions for AI use by state organs.
- AIL: Provides detailed provisions for specific AI applications, such as judicial AI, medical AI, social bots, biometric recognition, and autonomous driving.

Liability and Exemptions:

- MAIL: Offers liability exemptions for open-source AI providers under certain conditions and details about administrative penalties.
- AIL: Introduces safe harbor rules for service providers and includes punitive damages for severe violations.

Both the MAIL and AIL reflect China's proactive approach to AI governance, aiming to balance innovation with regulation and ethical considerations. While sharing common goals, they differ in regulatory structures, emphasis on licensing mechanisms, obligations for foundational models, and approaches to liability and international measures. Both proposals contribute valuable perspectives to the ongoing discourse on AI legislation in China. As China continues to refine its AI governance strategies, these expert contributions will possibly contribute to future legislation, shaping the country's role in the global AI landscape.

VII. CONCLUSION

China's regulatory approach to AI reflects its commitment to both advancing innovation and upholding rights and interests of state, society, and individuals in an increasingly complex digital landscape. Through a multifaceted and sector-specific framework of foundational laws, specific regulatory measures, sectoral guidelines, and technical standards, China seeks to balance the benefits of AI with the responsibilities it entails. In sum, China's AI regulatory landscape provides a robust, open, and very promising regulatory framework. As the global AI landscape evolves, China's approach will probably influence international norms and inspire continued discourse on achieving responsible, ethical, and sound AI development worldwide.

INTELIGENCIA ARTIFICIAL ¿A QUÉ OBJETIVOS DEBE RESPONDER LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN MÉXICO?

SUMARIO

I. Introducción **II.** Daños algorítmicos y la regulación basada en el marco de derechos humanos **III**. Sesgo y representación de grupos minoritarios en la inteligencia artificial **IV.** Custodia de datos y una posible política soberana de inteligencia artificial **V.** Conclusión **VI.** Bibliografía

> MARIEL GARCÍA MONTES

Investigadora tecno-social de medios y tecnología de la información, especialmente en cuanto a su contribución al interés público y el cambio social. Massachusetts Institute of Techonology (MIT)

> SAIPH SAVAGE

Profesora asistente en la Northeastern University, en el Khoury College of Computer Sciences, donde dirige el Northeastern Civic A.I. Lab. Northeastern University y Universidad Nacional Autónoma de México (UNAM).

I. INTRODUCCIÓN

En 2002, el Instituto Nacional de Estándares y Tecnología (NIST) de los Estados Unidos utilizó las fotografías de 120,000 mexicanos para entrenar su algoritmo de evaluación de tecnologías de reconocimiento facial (mejor conocido como Facial Recognition Vendor Test, o FRVT). Se trataba de personas que solicitaron una visa de entrada a los Estados Unidos, y acudieron a una cita consular para ser fotografiadas. En el momento del desarrollo del algoritmo, el Departamento de Estado concedió acceso a NIST a la base de datos con las fotografías de más de seis millones de mexicanos, y NIST tomó una muestra de los registros de personas que habían sido fotografiadas dos veces.[1]

A dos décadas y media del inicio de estos ejercicios de evaluación periódica, NIST se ha consolidado como la autoridad a nivel mundial en la evaluación de distintas métricas del reconocimiento y análisis facial. La importancia de estas evaluaciones es tal que distintos actores en la industria de la verificación computacional de la identidad hoy tratan los resultados de la evaluación de NIST como un sello de calidad, una garantía del buen funcionamiento de un sistema de reconocimiento facial. Una de las empresas que se benefician de su posición alta en en el ranking de NIST, que se lleva a cabo anualmente, es Incode Technologies. Se trata de una empresa de verificación biométrica fundada por Ricardo Amper, reconocida como historia de éxito de los emprendimientos mexicanos en Silicon Valley. Incode Technologies, que ha sido contratada

1 NIST. Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability. 13 de noviembre de 2002. https://www.nist.gov/system/files/documents/2021/10/25/nist_appendix_pact_nov02.pdf

por distintos actores públicos y privados en México, recalca en sus materiales publicitarios que la empresa recibe el primer lugar en su categoría en los rankings que genera NIST.[2]

El reconocimiento facial es una forma de visión computarizada que abstrae las caras humanas a objetos numéricos, permitiendo su análisis en grandes bases de datos, y es una de las formas de inteligencia artificial que han visto mayores avances tecnológicos en los últimos años. También es una de las más controversiales: su uso en el patrullaje predictivo y en procesos judiciales ha sido problematizado por sesgos raciales y otras prácticas discriminatorias[3] que surgen en las distintas capas del proceso de inteligencia artificial: la creación de datos para entrenamiento algorítmico, la construcción y estructura de las bases en las que se concentran los datos, el desarrollo de los algoritmos de aprendizaje y su aplicación en distintos ámbitos de la vida pública.

Una de las aplicaciones en la vida pública más frecuentes del reconocimiento y análisis facial computarizados en México es la seguridad pública. En abril de 2019, después de un viaje a China para hacer recorridos de las plantas de fabricación de Dahua Technology, el entonces gobernador de Coahuila Miguel Ángel Riquelme, anunció la compra de 1281 cámaras de videovigilancia con reconocimiento facial de esta compañía para su uso en 9 ciudades. La compra fue hecha por adjudicación directa, y el CEO global de Dahua Technology, Zhijie Li, estuvo presente en el anuncio en Saltillo[4]. Posteriormente, otras ciudades han hecho compras similares, tales como Ciudad Juárez en 2022[5]. Distintas organizaciones de la sociedad civil y medios de comunicación han denunciado la falta de transparencia sobre la efectividad de los algoritmos de reconocimiento facial que usan estas cámaras, y sobre los datos utilizados para su funcionamiento[6].

2 Incode. Incode's Stellar Performance in NIST FRTE 1:N Identification Testing: A Testament to Accuracy and Leadership in Facial Recognition. 17 de septiembre, 2024. https://incode.com/blog/incodes-stellar-performance-nist-frtr-1n-identification-testing/

3 Joy Buolamwini y Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." *Conference on Fairness, Accountability and Transparency*. PMLR, 2018.

4 Redacción de El Universal, "Preparan cámaras de vigilancia con reconocimiento facial en Coahuila." *El Universal*, 2 de abril de 2019, https://www.eluniversal.com.mx/estados/preparan-camaras-de-vigilancia-con-reconocimiento-facial-en-co ahuila/.

5 Elizabeth Rosales, "Ciudad Juárez compra equipos de videovigilancia a firmas chinas vetadas por EE.UU," *Empower LLC*, 7 de junio de 2023 https://empowerllc.net/2023/06/07/chihuahua-coahuila-videovigilancia/.

6 Federico Pérez Villoro y Paloma Robles, "Vigilancia biométrica: el tortuoso camino de Coahuila hacia el reconocimiento facial," *Quinto Elemento Lab*, 11 de noviembre de 2020, https://quintoelab.org/project/vigilancia-biometrica-reconocimiento-facial-coahuila.

Dahua Technology, la empresa subsidiada por el gobierno chino al centro de la controversia por las cámaras de reconocimiento facial en Coahuila, no es el único vendedor internacional de inteligencia artificial para la seguridad pública en México. También se encuentra el notable caso de Thales Group, la empresa francesa de tecnología de la información que opera el proyecto "Ciudad Segura" desde 2009. Este proyecto construyó cámaras y, más ampliamente, una red de infraestructura para las operaciones del C5 en la Ciudad de México[7] y en otras ciudades del país. En las palabras de Jean Marie Letort, entonces vicepresidente del consejo de ciberseguridad de Thales, su producto permite "no sólo vigilar y seguir la huella de algún criminal, sino que podremos prevenir delitos a través de sistemas de videovigilancia que usan inteligencia artificial con el fin de prevenir y evitar que algún delincuente culmine un delito"[8]. Hasta la fecha, Thales Group y los Centros de Comando y Control no han hecho públicos documentos que permitan evaluar los algoritmos con los que opera el análisis computarizado del video que capturan sus cámaras. No hay transparencia sobre el origen de los datos que se utiliza como *corpus* para tareas de entrenamiento algorítmico; no sabemos si hay procesos de toma de decisión automatizada, o en qué momentos el proceso requiere validación humana; tampoco sabemos de recursos de revisión para casos en los que el análisis computarizado dé resultados incorrectos.

Más allá del origen de los datos para la inteligencia artificial o de la creación de algoritmos y aplicaciones, o de la importación de algoritmos hechos en otros países, México es en la actualidad uno de los países que albergan los grandes centros donde se almacenan, procesan y distribuyen los datos al centro de la inteligencia artificial. De cierta manera, se puede decir que la nube se encuentra, entre otros sitios, en México. En mayo de 2024, Microsoft anunció el lanzamiento de su primer centro de datos a hiperescala en el país, con una proyección financiera de ganancias sobre los 70 mil millones de dólares[9]. Los

7 Madeleine Watterberger, "C5, la millonaria promesa de videovigilancia de CDMX que no previene ni resuelve delitos." *Animal Político*, 14 de febrero de 2021, https://www.animalpolitico.com/2021/02/c5-millonaria-promesa-videovigilancia-cdmx-no-resuelve-delitos.

8 Eréndira Reyes, "Cámaras con IA de la CDMX ya podrán evitar delitos", *Expansión*, 5 de junio de 2019, https://expansion.mx/tecnologia/2019/06/05/camaras-con-ia-de-la-cdmx-ya-podran-evitar-delitos.

9 Microsoft, "Microsoft launches its first hyper-scale cloud datacenter region in Mexico", *Microsoft*, 7 de mayo de 2024, https://news.microsoft.com/es-xl/microsoft-launches-its-first-hyper-scale-cloud-datacenter-region-in-mexic o/.

centros de datos requieren una cantidad amplia de insumos materiales para funcionar; especialmente electricidad y agua para enfriar la infraestructura. Microsoft construyó este nuevo centro de datos en Colón, un municipio del estado de Querétaro que, a inicios de 2024, estaba en estado de sequía extrema y excepcional[10].

Los ejemplos anteriores muestran que México ha servido como espacio de creación de datos originarios para el reconocimiento facial, y es probable que continúe siéndolo mediante contratos entre empresas internacionales y gobiernos locales. El país también ha sido espacio de implementación de distintos algoritmos de visión computerizada, tanto hechos en México como en otros países. Finalmente, el territorio mexicano también ha dado sustento material, en términos de energía y recursos naturales, para las infraestructuras necesarias para el procesamiento de datos y la ejecución de tareas de inteligencia artificial de compañías transnacionales.

Los hechos anteriores ejemplifican algunas de las distintas maneras y dimensiones en las que la inteligencia artificial ya ha repercutido en México, así como las maneras en las que México ha repercutido en la inteligencia artificial a nivel global. A partir de estos ejemplos, surgen las siguientes preguntas que permiten aterrizar las discusiones nacionales sobre reconocimiento facial y otras formas de inteligencia artificial a problemas no sólo en potencia, sino ya existentes, y que proponemos como guía para marcos regulatorios y líneas generales de implementación:

1. ¿Qué configuraciones institucionales permitieron que se usara las fotografías de personas mexicanas en Estados Unidos con propósitos de entrenamiento de algoritmos de reconocimiento facial en 2002, y qué salvaguardas debemos instituir en el país para proteger los datos biométricos de las y los mexicanos ante estos desarrollos tecnológicos, tanto en el territorio nacional como en otros países?
2. ¿Cómo podemos extender las protecciones que no tuvimos en 2002 a los usos actuales de datos biométricos en México, especialmente en el contexto de las contrataciones que los gobiernos locales hacen de equipo de identificación biométrica hechas en otros países? ¿Qué salvaguardas deben existir cuando haya convenios comerciales con empresas en países como China, Estados Unidos, Francia e Israel?

10 Viviana Estrella, "Querétaro, con 14 municipios en sequía extrema y excepcional," *El Economista*, 1 de enero de 2024, https://www.eleconomista.com.mx/estados/Queretaro-con-14-municipios-en-sequia-extrema-y-excepcion al-20240101-0050.html.

3. En el uso de la inteligencia artificial para tareas de seguridad pública, ¿cuáles son los aspectos de transparencia algorítmica que no pueden ser opaces en nombre de los secretos comerciales o de la seguridad nacional? ¿Qué regulaciones y provisiones de transparencia son necesarias para que en México podamos también llevar a cabo auditorías contra los sesgos discriminatorios en los datos, y para prevenir los daños algorítmicos?
4. ¿Qué elementos de análisis de impacto ambiental y de gestión de recursos hídricos deben formar parte del diseño y construcción de centros de datos en México? ¿Cómo podemos asegurar que las comunidades donde se construyan estos centros reciban los beneficios, y no sólo los daños, de la inteligencia artificial?

En este capítulo, analizamos algunas de las necesidades que se desprenden del panorama actual de la inteligencia artificial en México, así como algunas de las respuestas posibles que han surgido en la investigación de ciencias de la computación y el estudio crítico de datos.

I. DAÑOS ALGORÍTMICOS Y LA REGULACIÓN BASADA EN EL MARCO DE DERECHOS HUMANOS

CASO DE ESTUDIO:

***FanID* (Federación Mexicana de Fútbol e Incode Technologies)**

Los estudios sociales de la inteligencia artificial han generado análisis y evidencia de los costos sociales y los daños perpetrados por el uso de algoritmos en la toma de decisiones de la vida diaria. Por ejemplo, en 2016, el medio de periodismo de investigación ProPublica publicó una investigación sobre COMPAS, un algoritmo de predicción de reincidencia criminal ampliamente utilizado en la procuración de justicia en Estados Unidos. ProPublica analizó los casos de 10,000

personas acusadas de crimen y comparó las tasas de reincidencia predichas por COMPAS con las de reincidencia real dos años después. Los periodistas concluyeron que el algoritmo llevaba dos veces más a predicciones erróneas para las personas negras que para las personas blancas, entre otros sesgos raciales del algoritmo.[11] Este sesgo algorítmico se vio reflejado en las decisiones

11 Jeff Larson *et al*., "How We Analyzed the COMPAS Recidivism Algorithm", *ProPublica*, 23 de mayo de 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-

judiciales sobre fianzas y fechas de liberación. La investigación de ProPublica fue una de las primeras en demostrar los daños reales del uso de algoritmos predictivos para la toma de decisiones judiciales, e inspiró el análisis de los daños causados por algoritmos en otras industrias.

En México, a pesar de la existencia de una industria de ciencia de datos y generación de algoritmos predictivos, aún no contamos con instancias similares de evidencia sobre la efectividad de los algoritmos en uso, y menos aún de su impacto social. Se trata de algoritmos diseñados y empleados en contextos empresariales donde no existen requisitos de transparencia algorítmica. Ante esta opacidad comercial, la investigación social se ha enfocado en la oferta: la industria privada que vende ciencia de datos e inteligencia artificial como servicios de consultoría y desarrollo tecnológico. Sus clientes principales se encuentran en industrias como la banca, telemática y protección de transporte, y comercio. También ha habido desarrollos mexicanos en aplicaciones de la inteligencia artificial para la seguridad, la administración laboral y de recursos humanos, y la verificación de la identidad. La oferta de estos servicios se ha concentrado en aplicaciones específicas de la visión computarizada, tales como el reconocimiento facial y el análisis de datos biométricos.

Uno de los casos más emblemáticos del uso de la inteligencia artificial en tareas de seguridad en México es FanID, una verificación biométrica de la identidad de asistentes a estadios de fútbol. FanID surgió como solución tecnológica a raíz de un episodio de violencia extrema en un juego de fútbol en Querétaro. Ante críticas de la afición por la crisis de seguridad que dio pie a la tragedia del juego del 5 de marzo de 2022, la Federación Mexicana de Fútbol anunció un convenio con Incode Technologies para usar herramientas de inteligencia artificial y verificación biométrica para aumentar la seguridad en distintos estadios en el país.[12] La Secretaría de Seguridad y Protección Ciudadana anunció una serie de reuniones con la Federación Mexicana de Fútbol en las que discutió el proceso de diseño de FanID, pero negó los documentos de dichas reuniones cuando un ciudadano los solicitó a través de un recurso de acceso a la información[13].

recidivism-algorithm.

12 Incode Technologies. "How Fan ID Transformed Mexican League Football". *Incode Technologies*, 2 de marzo de 2023, https://incode.com/blog/how-fan-id-transformed-mexican-league-football/.

13 INAI. "Versión estenográfica de la Sesión Ordinaria del Pleno en materia de la Ley Federal de Transparencia y Acceso a la Información Pública (INAI)." *INAI*, 6 de septiembre de 2023, https://home.inai.org.mx/wp-content/documentos/AudiosyVersionesEstenograficas/06-09-2023_SesionPl eno_VE.pdf.

La promesa social y técnica que hace FanID es la eliminación del anonimato en los juegos de fútbol, dando a los equipos de seguridad herramientas para identificar y vetar a las personas que delinquen. En la práctica, FanID se construyó como una plataforma digital que incluye una aplicación móvil en la que los usuarios suben su fotografía, acompañada por su identificación oficial, y por otros datos personales como el número de teléfono y una dirección de correo electrónico. En la plataforma, esta aplicación se integra con una base de datos de datos personales. Los algoritmos de reconocimiento facial con los que Incode analiza estos datos permitirían que los equipos de seguridad pudieran identificar a las personas que incurren en agresiones en el estadio mediante fotografías, y señalarlas en la misma plataforma para evitar su ingreso a posteriores juegos.

Con la implementación de FanID, surgen distintas preguntas sobre los daños que podría causar un proyecto biométrico tan ambicioso. En diciembre de 2022, distintos estadios alrededor del país anunciaron el registro mediante FanID como requisito de entrada a ciertos juegos de fútbol. Por una parte, para su funcionamiento, FanID necesitaría participación universal de la afición que asiste a juegos de fútbol; por otra parte, ¿es realmente razonable condicionar la entrada a personas usuarias de celulares de alta gama, con la expectativa de cumplir una serie de tareas técnicas para registrarse? Se entiende que este requisito se volvería obligatorio en algún punto para personas mayores de edad.

Otra pregunta que surge en un espacio familiar tiene que ver con el manejo de datos de menores de edad. ¿Es suficiente el consentimiento parental al uso de sus datos biométricos un menor de edad, especialmente en etapas de infancia temprana, cuando acude al estadio con sus padres? ¿Qué efectos puede tener su inclusión en FanID hoy en su derecho a la privacidad en el futuro? Por el momento, para noviembre de 2024, el sitio de FanID anuncia que la verificación de menores de edad aún no comienza, pero que sí lo hará próximamente; actualmente se encuentra en revisión del Instituto Nacional de Acceso a la Información y Protección de Datos Personales (INAI).

Dado el panorama de filtración de bases masivas de datos personales en México, con la plataforma FanID surgen, en general, preguntas relacionadas con los daños posibles que podrían resultar de la filtración de los datos personales de la afición: las fotografías, las identificaciones oficiales, los números de teléfono. ¿Qué medidas deben tomar la Federación Mexicana de Fútbol e Incode Technologies para dar la mayor protección posible a estos datos personales? ¿Qué protocolos existen para mitigar los riesgos de posibles filtraciones masivas de información, así como los de violaciones individuales que pudieran surgir en el estadio o en los contextos de uso de la plataforma de FanID? Si FanID recopila

datos personales y sensibles como los datos biométricos, ¿podrán ser cruzados eventualmente con otras bases de datos? En 2013, el pleno del INAI discutió una solicitud de acceso a la información sobre cruces de la base de datos de FanID con Plataforma México, una plataforma de datos para la seguridad pública que utiliza la visión computarizada para encontrar información, por ejemplo, sobre personas con historial delictivo. Esta solicitud apunta a preocupaciones fundadas sobre los cruces de datos que permite en potencia la plataforma FanID, resaltando la necesidad de lineamientos de protección de datos que contemplen a FanID como parte de un contexto más amplio de uso de datos para la seguridad pública.

Algunas de estas preguntas se vieron reflejadas en los documentos sobre FanID del organismo garante del derecho a la protección de los datos personales. Después del anuncio del convenio entre la Federación Mexicana de Fútbol e Incode Technologies, con la participación de la Secretaría de Seguridad y Protección Ciudadana, el INAI atrajo el asunto para revisar una evaluación del impacto de FanID llevada a cabo por Incode Technologies y la Liga Mexicana de Fútbol en junio de 2022. Por la escala de la aplicación de este proyecto alrededor del país, se volvió necesario analizar el impacto que, en nombre de la seguridad pública, tendría en el manejo de los datos personales a nivel nacional.

En agosto de 2022, el INAI indicó la ejecución de un "análisis normativo-operativo respecto del tratamiento de datos personales que implicaría la implementación del FAN ID, con el objetivo de garantizar el debido cumplimiento de los principios, deberes y obligaciones previstas en la Ley Federal de Protección de Datos Personales en Posesión de los Particulares".[14] Una de sus recomendaciones fue hacer un análisis para corroborar, antes del diseño de la plataforma, que no se contara "con mecanismos menos invasivos para lograr la seguridad en los estadios" dados los antecedentes nacionales e internacionales del tratamiento de los datos biométricos a gran escala. Más tarde, en noviembre de 2022, el INAI advirtió que FanID "cuenta con brechas importantes para asegurar el tratamiento debido de datos personales que serán recabados".[15]

14 INAI. "Inai emite opinión técnica sobre el documento que presentaron la FMF y la LigaMX ante uso de sistema FanID en estadios". *INAI*, 17 de agosto de 2022. https://home.inai.org.mx/wp-content/documentos/SalaDePrensa/Comunicados/Comunicado%20INAI-248-22.pdf

15 INAI. "INAI concluye evaluación de Fan ID: FMF reforzará medidas para proteger datos personales y privacidad de aficionadas y aficionados". *INAI*, 6 de noviembre de 2022. https://home.inai.org.mx/wp-content/documentos/SalaDePrensa/Comunicados/Comunicado%20INAI-341-22.pdf

Considerando que en diciembre de 2022 los estadios comenzarían los anuncios del despliegue de FanID, surgen preguntas sobre la importancia que realmente se dio al análisis técnico del INAI. Por una parte, el involucramiento de una empresa de verificación de la identidad mediante biometría tuvo tal vez la consecuencia más importante. Significó que difícilmente se tomaría en cuenta la recomendación principal del INAI: la evaluación de mecanismos menos invasivos para cumplir las tareas de seguridad en los estadios de fútbol. Por otra, con el marco actual de protección de datos en México, los sujetos privados no tienen obligaciones de llevar a cabo evaluaciones de impacto en la privacidad, ni de seguir las recomendaciones del INAI resultantes de las evaluaciones hechas.

En otros marcos legales, o con configuraciones institucionales distintas, la plataforma FanID no hubiera podido continuar de la manera en que se diseñó y se usa actualmente. También habría documentos públicos sobre el análisis normativo-operativo que hizo el INAI; en la práctica, el Instituto negó las solicitudes de transparencia relacionadas con esta documentación, alegando que su dictamen contenía secretos industriales.[16] Es decir: el marco actual de protección de datos personales no protege a los ciudadanos de la recolección masiva de datos biométricos cuando es llevada a cabo por entes privados. Por secretos industriales y la Ley de Propiedad Industrial, tampoco obliga la publicación de documentos que permitan subsecuentes evaluaciones de impacto en la privacidad, y mucho menos la auditoría de sistemas de inteligencia artificial. Se trata de una configuración legal e institucional perfecta para promover la opacidad algorítmica.

Casos como el de FanID están, por lo tanto, al centro de las preocupaciones de la sociedad civil organizada y de otros sectores en el contexto de la inteligencia artificial en México y otros países en América Latina. Lo que este caso ejemplifica es la manera en que el avance tecnológico llega a contextos con marcos y valores que surgieron en épocas con posibilidades tecnológicas totalmente distintas a las de hoy día. En afán de aprovechar las nuevas posibilidades para buscar soluciones a problemas antiguos, se apresura el diseño de productos, sin antes examinar el legado sociotécnico que dejaron intentos previos de solución y gobernanza. Las tecnologías de visión computarizada para la identificación biométrica llegan a un contexto que da mayores protecciones a la propiedad industrial y a los intereses privados que a los derechos individuales y colectivos a la privacidad.

16 Red en Defensa de los Derechos Digitales. "INAI y LigaMX ocultan funcionamiento del FanID." https://r3d.mx/2023/01/10/inai-y-ligamx-ocultan-funcionamiento-del-fan-id/

Es por eso que distintas asociaciones profesionales, organizaciones de la sociedad civil y personas expertas en inteligencia artificial responsable convocan al desarrollo de un marco de diseño y regulación de la IA que se base en los derechos humanos.

> "Debemos apuntar a que todo intento de regular una tecnología se centre en la protección de los derechos humanos como principal e inclaudicable objetivo, teniendo particular consideración por la realidad social existente en el territorio donde impactará dicha regulación, para, al mismo tiempo y con la misma relevancia, considerar los mecanismos necesarios que fomenten la innovación y el desarrollo de competencias técnicas en los territorios."
>
> – Access Now, *Radiografía normativa:* ¿Dónde, *qué y cómo se está regulando la inteligencia artificial en América Latina*?, 2024.

Hidvegi y Leufer proponen centrar la discusión de la inteligencia artificial en derechos humanos, reforzando los derechos básicos de las personas que serían mayormente afectadas por aplicaciones de la inteligencia artificial.[17] Es decir: toda aplicación de la inteligencia artificial debe examinar primero el impacto en los derechos básicos de todas las personas, y abandonar los proyectos que sacrifiquen los derechos de algunos por el beneficio de otros. Se trata de una aplicación sociotécnica de una máxima del derecho: los derechos humanos no son negociables.

Hidvegi y Leufer, así como organizaciones de la sociedad civil, proponen el enfoque de derechos humanos en discusiones sobre inteligencia artificial como contraparte al modo de análisis que sacrifica los derechos básicos de grupos minoritarios en pos de proteger los intereses de grupos más amplios. El enfoque utilitarista que subyace a esta toma de decisiones ha llevado a la aceptación de tecnologías y prácticas que la sociedad difícilmente aceptaría si analizara el impacto en los derechos de poblaciones enteras.

Desafortunadamente, la regulación de la IA en México ha seguido pautas incompatibles con el enfoque de derechos humanos. Tomar como referencia el modelo regulatorio de la Unión Europea ha permitido la adopción de aplicaciones de la IA alineadas con principios de seguridad pública y nacional, a pesar de las violaciones de derechos humanos que podrían perpetrar. Un estudio

17 Fanny Hidvegi y Daniel Leufer, "The EU should regulate AI on the basis of rights, not risks". EDRI, 24 de marzo de 2021, https://edri.org/our-work/eu-should-regulate-ai-on-the-basis-of-rights-not-risks/.

reciente de los caminos regulatorios para la IA en América Latina concluye que la regulación en México "se centra en riesgos que aún no existen y quita la mirada de temas específicos como la vigilancia biométrica masiva bajo premisas de "seguridad nacional".[18] La plataforma FanID es un ejemplo de la vigilancia biométrica masiva a la que alude este reporte; el marco europeo permite el uso de plataformas similares para el control migratorio en la región, y especialmente en los campos de refugio.

"Un modelo centrado en derechos humanos ayudará a delinear cada una de las obligaciones en materia de derechos humanos que deben cumplirse en cada paso del ciclo de vida de la IA."[19]

– Al Sur, "Los caminos regulatorios para la IA en América Latina", 2024

¿Con qué elementos debe contar un marco regulatorio que centra los derechos humanos en discusiones de IA? Access Now propone la adopción de principios de transparencia, instrumentos de supervisión eficientes que incluyan estudios de impacto en derechos, cumplimiento de la ley (y la capacidad de aplicar sanciones), y análisis localizado. El caso de FanID cumple parcialmente con sólo una de estas recomendaciones: la evaluación de impacto en el derecho a la privacidad.

II. SESGO Y REPRESENTACIÓN DE GRUPOS MINORITARIOS EN LA INTELIGENCIA ARTIFICIAL

CASO DE ESTUDIO:

Cámaras de seguridad pública con reconocimiento facial en Coahuila y Ciudad Juárez

En 2017, Joy Buolamwini, científica de la computación en el Instituto Tecnológico de Massachusetts, decidió estudiar el funcionamiento de los algoritmos de reconocimiento facial. Buolamwini, como investigadora negra, pronto se dio cuenta de que los sistemas constantemente presentaban fallos técnicos cuando ella intentaba usar algoritmos para reconocer su propia cara.[20] Para lograr que

18 Consorcio Al Sur, "Los caminos regulatorios para la IA en América Latina". 2024 https://www.alsur.lat/sites/default/files/2024-09/ALSUR%20-%20IA%20en%20Latam%20%5BESP%5D.p df

19 *Idem*.

20 Joy Buolamwini y Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." *Conference on Fairness, Accountability and Transparency*. PMLR, 2018.

su cara fuera legible ante los algoritmos y cámaras de sus pruebas, a veces tenía que ponerse un antifaz blanco. Ante este sorprendente hallazgo, se ha dedicado desde entonces a estudiar las maneras en que la inteligencia artificial excluye y perpetra daños contra grupos minoritarios, como lo son las personas negras en Estados Unidos.

La investigación de Buolamwini catalizó una ola de auditorías e investigaciones sobre las maneras en que la inteligencia artificial incurre en discriminación racial. Tarleton Gillespie describe estas consecuencias como "daños representacionales" de la inteligencia artificial, los cuales articulan los daños perpetrados por sistemas al incluir pocos o malos datos de una población o grupo minoritario. Es decir: los algoritmos de reconocimiento facial no leían la cara de Buolamwini porque no habían sido entrenados con bases de datos de caras negras. La discusión resultante de esta línea de investigación se conoce más popularmente como una discusión sobre los sesgos de la inteligencia artificial.

Tal como en el caso de estudio anterior sobre daños algorítmicos, en México aún no tenemos proyectos de investigación amplios que demuestren y permitan la comprensión de las maneras en que el sesgo algorítmico ocurre en territorio mexicano. Sin embargo, algunos desarrollos de la IA en los últimos años en México nos dejan con preguntas que apuntan en esa dirección. Se trata del mismo tipo de tecnologías y algoritmos que no reconocieron la cara de Buolamwini en las pruebas que ella hizo, y que ahora son utilizados alrededor de México para el cumplimiento de tareas de seguridad pública.

En abril de 2019, como ya se indicó, después de un viaje a China, el entonces gobernador del estado de Coahuila, Miguel Ángel Riquelme, anunció la compra de 1281 cámaras equipadas con análisis computarizado. Las cámaras serían adquiridas por adjudicación directa de la compañía Dahua Technology, y serían utilizadas en 370 localidades en 9 ciudades.[21] Otras ciudades del país han seguido la misma trayectoria. Por ejemplo, también en 2019 anunció el Gobierno de la Ciudad de México sus planes de renovar y aumentar 18,500 cámaras de videovigilancia en la ciudad para dotarlas con alta resolución, visión 360°, wifi gratis, mayor capacidad de almacenamiento (de 7 a 30 días), así como reconocimiento facial y de placas. Ciudades en Chihuahua y en Jalisco también hicieron adquisiciones similares.

21 "Segundo informe de gobierno, Miguel Ángel Riquelme Solís, Gobernador Constitucional del Estado de Coahuila de Zaragoza, 2017-2023" [Second mandate report, Miguel Ángel Riquelme Solís, Constitutional Governor of the State of Coahuila of Zaragoza], Gobierno de Coahuila, accessed 5 February 2023, https://coahuila.gob.mx/archivos/pdf/micrositio/Documentos%20del%20Segundo%20Informe%20de%20 Gobierno/2o-Informe.pdf

La llegada a México de tecnología de reconocimiento facial hecha en China muestra paralelismos con la entrada de las mismas tecnologías a África. En el mismo año, Zimbabwe celebró un acuerdo comercial con empresas chinas, tales como CloudWalk e Hikvision, para ser el primer país en introducir sus productos al continente.[22] Este programa resultó ser un piloto para expandir las capacidades técnicas de estas empresas para reconocer caras africanas; donaron equipo a Zimbabwe, y a cambio, este entregó los datos de sus ciudadanos.[23] En el caso de Coahuila, dos circunstancias no permitieron saber si se prometió el mismo intercambio que en Zimbabwe: la decisión del gobierno contratante de reservar los contratos y todos los documentos de la adquisición, y la opacidad del secreto industrial de Dahua Technology. No hay documentos públicos que indiquen cómo los datos de las caras mexicanas son almacenados o transferidos, o qué usos les puede dar Dahua Technology.

Uno de los aspectos del funcionamiento de estas tecnologías que es difícil de conocer en el contexto de opacidad institucional tiene que ver con las fuentes de datos para el entrenamiento de algoritmos, y para cruces de información. ¿Cuál es el *corpus* que utilizan para funcionar en el mercado mexicano, y cuál es la proveniencia de estos datos? Sabemos que, por ley, la base de datos biométricos más amplia del país, que corresponde al Registro Federal Electoral, no puede ser utilizada con propósitos de entrenamiento algorítmico ni de cruce de información para cámaras de seguridad. Sin embargo, hay otras bases de datos que podrían estar en uso. ¿Las cámaras de seguridad en Coahuila hacen cruces con Plataforma México, por ejemplo? ¿O acaso con las bases de datos de la Secretaría de Relaciones Exteriores?

Y, más allá de las fuentes de datos y del entrenamiento algorítmico, no hay claridad sobre cómo se va a utilizar la información resultante de estas tecnologías. ¿Hay procesos de validación humana cada vez que la tecnología arroja un *match* o resultado que indique la identidad de una persona? ¿Va a haber requisitos de revisión si la tecnología llega a cometer errores en la clasificación de las personas? Si las tecnologías están integradas a través de los sistemas de seguridad pública, ¿los datos resultantes van a ser utilizados también en procesos judiciales?

22 The Herald, "Chinese tech revolution comes to Zim". The Herald, 9 de octubre de 2019. https://www.herald.co.zw/chinese-tech-revolution-comes-to-zim/

23 Jili Bulelani, "The Rise of Chinese Surveillance Technology in Africa". Electronic Privacy Information Center. 31 de mayo de 2022. https://epic.org/the-rise-of-chinese-surveillance-technology-in-africa/

En general, este despliegue de cámaras de vigilancia con capacidades de análisis computarizado en el espacio público se da de manera paralela a otras instancias de inteligencia artificial en la vida pública. Por ejemplo, en la última década, empresas privadas han desarrollado algoritmos predictivos para sectores como la banca y la compra minorista en México. Estos usos de la inteligencia artificial en la vida pública nos hacen preguntarnos: ¿Qué sesgos de datos existen en las aplicaciones de la IA implementadas hoy en día? ¿Qué necesitaríamos para confiar en que estas aplicaciones no van a reforzar los sesgos sociales más dañinos en sus operaciones? ¿Quiénes son las personas en las que recaen los daños de estas prácticas más frecuentemente?

"Para validar efectivamente estas tecnologías hay que ver dónde se equivocan, cómo se equivocan y cómo esos errores implican o generan un efecto en lo que realmente queremos lograr".[24]

Los sesgos algorítmicos han dado pie a la necesidad de apoyar iniciativas de representación de grupos, contenidos, y valores minoritarios en las distintas capas que constituyen la inteligencia artificial. Algunos esfuerzos de investigación han promovido esta representación en la inteligencia artificial de grupos minoritarios a través del diseño de herramientas que intencionalmente incorporan datos subrepresentados en este campo. Estos esfuerzos requieren, en algunas instancias, la creación de los datos con los cuales se entrena los modelos de aprendizaje computarizado. Por ejemplo, el proyecto de Datos contra el Feminicidio del Laboratorio de Feminismo de Datos en el Instituto Tecnológico de Massachusetts tuvo que generar sus propios cuerpos de datos para desarrollar una herramienta de detección de noticias sobre el violencia de género, dada la importancia que éstas tienen para los esfuerzos regionales de creación de registros nacionales del feminicidio.[25]

Podemos encontrar más ejemplos de este tipo de iniciativas en los proyectos que apoya la Red de Investigación Feminista de Inteligencia Artificial, que busca identificar y corregir prejuicios en la inteligencia artificial y otros espacios digitales, especialmente para resarcir el sesgo de género. En Argentina, surge una aplicación contra el sesgo de género en el sistema judicial: AymurAI, una iniciativa de la organización Data Género y la Corte Criminal 10 para apoyar la recolección de datos sobre violencia de género en sentencias judiciales. El uso

24 Carlos Duchanoy, Centro de Investigación en Computación del Instituto Politécnico Nacional, en entrevista con *Quinto Elemento Lab*.

25 Catherine D'Ignazio, I. Cruxên, H. Suárez Val, A. Martínez Cuba, M. García-Montes, S. Fumega y W. So. "Feminicide and counterdata production: Activist efforts to monitor and challenge gender-related violence". *Patterns* 3(7), 2022.

de la inteligencia artificial en este contexto permite agilizar la recolección de datos para propósitos de investigación sobre igualdad, pues, de otra manera, los datos tendrían que ser consultados manualmente en cada una de las sentencias, con grandes costos de tiempo, trabajo e impacto emocional que ya han sido documentados en proyectos de creación de datos contra la violencia de género[26]. Los datos de violencia de género en el sistema judicial han sido centrales para las discusiones sobre temas como el feminicidio.

Otro ejemplo de herramienta de la inteligencia artificial contra el sesgo de género, también en Argentina, es el de E.D.I.A. Este proyecto es una colaboración entre la organización Vía Libre y dos profesoras de ciencias de la computación, que desarrolla una herramienta de análisis de software de procesamiento del lenguaje natural para buscar sesgos de género en textos. En su piloto, probaron la herramienta con 500 trabajadores de la educación media superior en Argentina para que puedan incorporar la herramienta en sus lecciones escolares. Se trata de un uso de la inteligencia artificial para poner en evidencia las maneras en las que el sesgo de género se materializan en la vida diaria.

Finalmente, cabe resaltar el trabajo que se ha hecho para aprovechar la inteligencia artificial para la defensa de los derechos de las y los trabajadores. En la actual economía de plataforma, donde trabajadoras independientes se registran para proveer distintos servicios, el diseño de las plataformas imposibilita la convivencia entre trabajadoras.[27] La Independiente es otro proyecto apoyado por la Red de Investigación Feminista que busca utilizar la inteligencia artificial para promover las redes de comunicación y apoyo entre trabajadoras que no se pueden dar en las plataformas donde ellas trabajan.[28]

Estos proyectos de inteligencia artificial, así como otros apoyados por la Red de Investigación Feminista, ejemplifican la manera en que se puede promover el desarrollo de inteligencia artificial que incorpore y centre a los grupos que históricamente han estado ausentes de los procesos de desarrollo computacional. Para reducir los sesgos de género, es necesario incorporar a

26 *Ibid*.

27 Mary Gray, Mary y Siddharth Suri. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books, 2019.

28 Maya De Los Santos, Norma Elva Chávez, Alberto Navarrete, Cristina Martínez Pinto, Luz Elena González, Tatiana Telles-Calderon, and Saiph Savage. 2023. La Independiente: Designing Ubiquitous Systems for Latin American and Caribbean Women Crowdworkers. In Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), October 08--12, 2023, Cancun, Quintana Roo, Mexico.

las mujeres y a todas las personas de la diversidad sexogenérica, y buscar las aplicaciones de la tecnología que puedan apoyar en la lucha por la igualdad.

III. CUSTODIA DE DATOS Y UNA POSIBLE POLÍTICA SOBERANA DE INTELIGENCIA ARTIFICIAL

CASO DE ESTUDIO:

CBP One, aplicación de administración migratoria en EEUU

Al inicio de este capítulo, resaltamos el caso de la base de datos consulares mexicanos que sirvió como cuerpo de entrenamiento para la evaluación de algoritmos de reconocimiento facial. En 2002, no hubo contrapesos que impidieran que el Departamento del Estado de Estados Unidos diera a NIST acceso a una base de datos de más de seis millones de mexicanos. Es difícil imaginar que las personas que se tomaron una fotografía para pedir una visa de entrada en 2002 se imaginarían en ese momento que su cara entrenaría tecnología computacional para la detección de caras, especialmente antes de la prominencia del reconocimiento facial en el discurso público.

Esta misma dinámica de reapropiación de datos personales con propósitos distintos de análisis computarizado aparece también en el caso del Fan ID de Incode Technologies y la Federación Mexicana del Fútbol. Es difícil imaginar que una persona que quiere ir a un partido de fútbol también está interesada en consentir el uso de sus datos personales para análisis automatizados hechos por una compañía privada, y el marco legislativo hoy no nos permite saber cuáles son exactamente los usos que acordaron privadamente Incode Technologies y la Federación contratante.

Finalmente, esta dinámica se vuelve evidente con las contrataciones de cámaras de seguridad pública con capacidades de reconocimiento facial, especialmente las creadas en China. Las experiencias internacionales muestran evidencia de la extracción de datos que permiten estos acuerdos comerciales, y es difícil imaginar, en el contexto de opacidad contractual, que los gobiernos locales de México habrían abogado por los derechos a la privacidad de las personas que son videograbadas con estas cámaras. ¿Cómo podemos garantizar el derecho a la privacidad de las personas mexicanas cuando sus datos personales cruzan fronteras, entregados por los mismos gobiernos en turno?

Los daños de esta trayectoria de extracción se vuelven más evidentes cuando surgen tecnologías que aprovechan datos mexicanos para después

utilizarlos contra los intereses de las personas en el país. Un ejemplo paradigmático de esta relación es el de CBP One, una aplicación móvil de manejo migratorio que lanzó el Departamento de Aduanas y Protección Fronteriza (CBP, por sus siglas en inglés) de Estados Unidos en octubre de 2020, en el contexto de la contingencia global por la pandemia de COVID-19. Esta aplicación es la única plataforma mediante la cual los migrantes que llegan a la frontera entre Estados Unidos y México pueden solicitar asilo. Como parte de la plataforma, los migrantes de algunos países en América Latina tienen que dar sus datos biométricos. A pesar de que esta aplicación es administrada en los Estados Unidos, las personas la usan en territorio mexicano, y muchas de ellas son también ciudadanas de México.

Ante el impacto de CBP One en los derechos de personas migrantes, y especialmente en sus derechos a la privacidad y la protección de datos personales, organizaciones internacionales han criticado el uso de esta aplicación en el contexto migratorio. Amnistía Internacional critica la promesa tecnológica que hizo CBP One, garantizando que sería más fácil para las personas solicitar asilo al pedir una cita a través de una aplicación, mostrando que la realidad resultó ser más complicada. En un estudio con cientos de personas migrantes y de las organizaciones que les acompañan, Amnistía Internacional encontró que enfrentaron dificultades significativas para usar la aplicación.[29]

Distintos grupos de investigación han abordado las características técnicas de CBP One, documentando las decisiones tecnológicas con mayor impacto para la privacidad de las personas migrantes. Algunos ejemplos son el requisito de activar la geolocalización de un celular para el uso de CBP One, y la obligación de mandar fotografías del usuario tomadas a través de la aplicación. Una decisión tecnológica que también ha resultado controversial es la prueba de "vivacidad" (*liveness*) que hace la aplicación a partir de las fotografías enviadas, dado que organizaciones de los derechos de migrantes de países como Haití y África han reportado que la aplicación no reconoce fotos de personas con tonos de piel más oscuros[30].

La ex relatora especial de la ONU sobre las formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia, E. Tendayi Achiume, expresó su preocupación por el potencial de CBP One de

29 Amnistía Internacional. "CBP One: ¿Una bendición o una trampa?" *Amnistía Internacional*, 8 de mayo de 2024. https://www.amnesty.org/es/documents/amr51/7985/2024/es/

30 American Immigration Council. "CBP One: An Overview". *American Immigration Council*, 2 de junio de 2023. https://www.americanimmigrationcouncil.org/research/cbp-one-overview

crear problemas de privacidad y no discriminación[31]. Ella propone entender tecnologías como CBP One como mecanismos de las fronteras raciales digitales, utilizados para dificultar la movilidad y migración de las personas según su raza.[32] La discriminación contra caras negras que perpetra la aplicación es el ejemplo más evidente de esta tecnología para la exclusión; sin embargo, el tipo de datos recolectados permite daños de distintos tipos, y se vuelve apremiante en el contexto de emergencia que viven las y los migrantes en América Latina hacer una evaluación de los impactos de tecnologías como CBP One.

> "A pesar de que la digitalización de la frontera parecería tener un abordaje más humano que el de las tecnologías tradicionales de manejo fronterizo, un estudio de la frontera sur [de los Estados Unidos] encontró que la frontera "inteligente" resultó en más del doble de muertes fronterizas. Quienes están muriendo son casi exclusivamente migrantes y refugiados negros y de color."[33]

Como en todos los casos anteriores, la documentación sobre el desarrollo técnico de la plataforma, de sus algoritmos de inteligencia artificial, y de sus evaluaciones de impacto no se encuentra abierta al público. La organización internacional Access Now y la Clínica de Ciberderecho de la Universidad de Harvard comenzaron una serie de solicitudes de acceso a la información sobre este programa en Estados Unidos, las cuales culminaron en un juicio contra CBP, que Access Now y Harvard ganaron. Esto llevó a la publicación de casi 3000 páginas de documentación sobre CBP One[34], que éstas y otras organizaciones están analizando por sus consecuencias para los derechos migrantes.

---

31 Naciones Unidas. "Discriminación racial y xenófoba y uso de tecnologías digitales en el control de fronteras e inmigración. Informe de la Relatora Especial sobre las formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia, E. Tendayi Achiume." *A/HRC/48/76*, 17 diciembre de 2021. https://documents.un.org/doc/undoc/gen/g21/379/64/pdf/g2137964.pdf

32 E. Tendayi Achiume. "Digital Racial Borders." *AJIL Unbound*. 115:333-338. doi:10.1017/aju.2021.52

33 E. Tendayi Achiumi, *Digital Racial Borders*.

34 Access Now. "Agencia estadounidense CBP publica registros de la polémica aplicación CBP One tras demanda de Access Now". *Access Now*, 10 de octubre de 2024, https://www.accessnow.org/press-release/u-s-cbp-releases-records-on-controversial-cbp-one-app-after-a ccess-nows-lawsuit-es/

Particularmente con el caso de CBP One, donde un gobierno extranjero recolecta los datos personales de personas mexicanas en territorio mexicano, surgen preguntas sobre la custodia de los datos en juego. ¿Qué marco regulatorio tendría que existir para proteger el derecho a la privacidad de las personas mexicanas ante estas aplicaciones de la inteligencia artificial? ¿Qué medidas se tendría que tomar para garantizar que los datos recolectados en México no sean utilizados en contra de personas mexicanas por gobiernos o por empresas en el extranjero? ¿Cómo podríamos extender la misma garantía del derecho a la privacidad a las personas migrantes y refugiadas de otros países que cruzan el territorio mexicano?

Estas preguntas apuntan a la necesidad de una política de soberanía de datos en México. La soberanía de datos se vuelve indispensable en un contexto en el que cambia el valor de los datos analógicos con la medida de los avances en el análisis computacional. Las fotografías que las personas entregaban en sus solicitudes de visa de entrada cambiaron de valor cuando se volvió posible entrenar algoritmos de reconocimiento facial con ellas. La fotocopia y la fotografía de respaldo de una identificación oficial, como la credencial de elector, cambiaron de valor cuando empresas privadas las comenzaron a almacenar en bases de datos de asistentes de fútbol. La *selfie* tomada con un celular cambió de valor cuando fue sujeta a una prueba de *vivacidad*, intencional o accidentalmente diseñada para excluir a las caras negras, en una base de datos fronteriza.

Para idear una política de soberanía de datos en el país, debemos reconocer la importancia que tienen los mismos datos para el desarrollo de políticas públicas y para la promoción del bienestar social.[35] Si socialmente vamos a crear datos para el funcionamiento de nuevas herramientas para solucionar los problemas que enfrentamos, debemos consecuentemente aceptar la importancia y la sensibilidad de estos datos. Los procesos de creación y uso de datos deben estar acompañados por regulaciones claras que garanticen sólo los usos para el bienestar, y protejan a los mexicanos de la explotación que pueden hacer los intereses privados y gubernamentales locales y extranjeros.

> "Concebimos una política soberana como aquella que se define a partir de los intereses de la propia sociedad en la que emerge, sin injerencia de los intereses extranjeros, que responde a las prioridades y necesidades diversas

35 Paola Ricaurte (ed). *Hacia una política de Inteligencia Artificial centrada en los derechos humanos: Oportunidades y Recomendaciones*. 7 de octubre de 2024, https://zenodo.org/records/13899856.

de México. Además, una política anclada en una perspectiva de derechos humanos busca actuar proactivamente para que se reduzcan las brechas de la desigualdad y que todas las personas, sin importar su condición, vean sus derechos protegidos frente a los riesgos de los sistemas basados en la IA y, de así desearlo, puedan tener acceso a los beneficios de estas tecnologías para alcanzar una vida digna."[36]

IV. CONCLUSIÓN

En este capítulo, las autoras analizamos algunos de los desarrollos y aplicaciones de la inteligencia artificial que ya han tenido lugar en México, así como las tensiones que revelan entre el análisis computarizado y el bienestar social. Nos enfocamos principalmente en las tecnologías de reconocimiento facial, un tipo de visión computarizada que prolifera a gran velocidad en el país en procesos de identificación, seguridad pública y privada, y control migratorio. Nuestro análisis explica algunos de los requisitos técnicos con consecuencias para los derechos de las y los mexicanos, así como los obstáculos que existen actualmente para las evaluaciones de impacto y la determinación de responsabilidad algorítmica. Especialmente describimos las dificultades para la investigación y la auditoría de la inteligencia artificial en el contexto de opacidad que permiten las configuraciones de actores públicos y privados, tanto domésticos como extranjeros en México. Mostramos cómo estas decisiones acarrean daños algorítmicos no sólo en potencia, sino ya vistos en la vida pública del país. A partir de este análisis, las autoras encontramos oportunidades para el marco regulatorio de la inteligencia artificial en México:

1. Para prevenir y resarcir los daños algorítmicos de las aplicaciones futuras y actuales de la inteligencia artificial en México, el marco regulatorio y las líneas generales de implementación deben centrar los derechos humanos. Ante todo desarrollo o aplicación, debemos hacer la pregunta: ¿A qué personas afectaría más este sistema, y cómo podemos garantizar que el sistema no violará sus derechos básicos?

36 Paola Ricaurte *et al.*, "Hacia una política de Inteligencia Artificial centrada en los derechos humanos: Oportunidades y Recomendaciones".

2. Para evitar los daños algorítmicos que tienen que ver específicamente con sesgo y prejuicio social, México debe invertir en los desarrollos de los grupos minoritarios y poblaciones ausentes de la inteligencia artificial que buscan generar sus propios datos y tecnologías. La ciencia de datos feminista interseccional es un ejemplo de los marcos conceptuales y prácticas que pueden guiar estos esfuerzos.
3. Ante el cambio en el valor de los datos personales en el contexto de las tecnologías de inteligencia artificial, uno de los mayores enemigos al derecho a la privacidad en México es la extracción de datos personales mexicanos para su capitalización en el extranjero. Para enfrentar esta realidad, la política pública de inteligencia artificial en el país debe dirigirse hacia la soberanía de los datos mexicanos.

LA INTELIGENCIA ARTIFICIAL GENERATIVA Y SU IMPACTO EN LA CREATIVIDAD Y EN EL DERECHO INTELECTUAL

SUMARIO

I. El Derecho como ciencia reguladora **II.** La Inteligencia Artificial **III.** Panorama Internacional **IV.** Panorama en México **V.** Propuestas **VI.** Bibliografía

> KIYOSHI TSURU

Socio fundador y director de TMI Law. Kiyoshi es un abogado líder especializado en propiedad intelectual, particularmente en los campos de la innovación y el ciberespacio, resolución de disputas sobre nombres de dominio, protección de derechos de autor en el entorno digital, protección de software, patentes, indicaciones geográficas, transacciones electrónicas relacionadas con los derechos de propiedad intelectual y arbitraje.
Country Manager de BSA | The Software Alliance.
Representante en México de la Entertainment Software Association (ESA).

I. EL DERECHO COMO CIENCIA REGULADORA

Según el Doctor José Ramón Cossío Díaz, el Derecho no es una ciencia absoluta, independiente o autosuficiente. Existe una relación simbiótica entre Derecho y Ciencia, pues el primero intenta regular a la segunda, pero debido a la exponencial evolución científica, éste necesita nutrirse de aquella y entender que, de no solicitar el auxilio de la esfera científica, el Derecho se encuentra limitado, impedido para comprender y resolver las controversias surgidas en nuestra compleja realidad contemporánea[1].

Por su parte, el Doctor Villoro Toranzo retoma la noción de Santo Tomás de Aquino de que "la Ley no es más que un ordenamiento de la razón, en orden al bien común, promulgado por aquel que tiene a su cuidado a la comunidad" para explicar la esencia normativa del Derecho y la función contemplativa del jurista. En específico, dicho tratadista refiere que la labor del jurista respecto de la creación de una norma radica en retomar los conocimientos de otras ciencias (como la psicología y la economía) para conocer la realidad jurídica. Una vez conocida la realidad jurídica, la labor del jurista es ordenarla a través de normas, para lo cual primero debe valorar si los datos jurídicos (que componen la realidad jurídica) son justos o injustos, después construir el instrumento que regulará dicha realidad y versará sobre dichos datos jurídicos para finalmente aplicar el referido instrumento[2].

1 Cossío Díaz José Ramón, *Derecho y Ciencia*, Editorial Tirant lo Blanch, Ciudad de México, México, 2015, Capítulo I.

2 M. Villoro Toranzo, *Introducción al estudio del Derecho*, Editorial Porrúa, Ciudad de México, México, 1972.

Se ha vuelto tan generalizado el planteamiento de Villoro Toranzo y de otros autores en cuanto a que el proceso de creación del Derecho entendido como norma, conlleva un ejercicio contemplativo de la realidad para entenderla y posteriormente regularla conforme a los cánones de lo justo y lo injusto según un determinado contexto social, político y temporal, que a veces parecemos olvidarnos de la noción de que el Derecho es inherentemente reactivo, en el sentido de que, primero debe de existir el objeto de una norma para posteriormente comprenderlo y luego intentar normarlo. Lo anterior toma una mucho mayor relevancia cuando pensamos en lo vertiginosamente que se transforman nuestros contextos sociales y culturales en la actualidad. Este principio, imprevisible para los autores que han teorizado al respecto, ha sido rebasado por el gran ímpetu y velocidad en los cambios del estado de la técnica en el ecosistema tecnológico. Como se ha establecido previamente, podemos concluir que, al visualizar la realidad y el Derecho, es claro que el segundo teleológica y naturalmente va después que la primera.

El desarrollo tecnológico siempre ha tenido como consecuencia que la realidad jurídica a la que se han referido Cossío Díaz y Villoro Toranzo sea aún más difícil de entender y contemplar para posteriormente normar. La realidad jurídica acelerada por el desarrollo tecnológico cambia más rápido que la capacidad que tiene el Derecho para digerir los datos jurídicos y tecnológicos y crear supuestos y consecuencias normativas que le recaigan. Esto explica la rapidez con la que los ordenamientos jurídicos se vuelven anacrónicamente inaplicables u obsoletos y más recientemente, los ordenamientos jurídicos que versan sobre una innovación específica y se "quedan cortos" ante la acelerada evolución de la tecnología.

Siguiendo con esta esquematización de la realidad fáctica y el Derecho como actividad reguladora, se hace evidente que mientras la sociedad fue creando nuevos contextos sociales, el Derecho fue respondiendo con normas que intentaran prevenir y resolver disputas en su entorno contemporáneo. Por ejemplo, en etapas tempranas de nuestra civilización, la protección y preservación de la vida y el respeto a la propiedad fueron primordiales. Con el paso del tiempo, la sociedad requirió de principios que garantizaran la libertad creadora del alma humana y resguardaran al mayor exponente de la libertad de expresión: el derecho a que se le reconozca a un creador su carácter de autor, se respete la integridad de su obra, y se garantice su facultad personalísima e inalienable de autorizar o prohibir su explotación y de percibir los frutos de ésta.

Uno de los primeros antecedentes registrados en nuestra historia occidental es el sistema de protección de recetas de Síbaris, en la antigua Grecia, documentado en una instancia inicial por Ateneo de Náucratis y posteriormente

retomado por el historiador alemán Conrad Cichorius en 1922. Este sistema consistía en que los cocineros que crearan un platillo nuevo, que resultara favorito del paladar de los comensales, recibían el derecho exclusivo a prepararlo con exclusión del resto de los pobladores, por el término de un año. Esta protección estaba circunscrita al territorio de Síbaris[3].

Otro antecedente, considerado por algunos el nacimiento del Derecho Intelectual[4] es el Estatuto de la Reina Ana de 1710, que sustituyó la "*Licensing Act*", y que estableció por primera vez que el derecho a realizar una copia (*copyright*) le pertenecía a más personas además de los editores de libros (*stationers*), otorgando derechos sobre la impresión de una obra determinada, durante 14 años que podían renovarse por otros 14 años si el autor seguía con vida, o bien durante 21 años no renovables para aquellas obras anteriores a la entrada en vigor del Estatuto. Autores como Deborah Halbert, con quienes concuerdo, han sentenciado que la verdadera intención subyacente en el Estatuto de la Reina Ana de 1710, más que constituir un marco jurídico de protección a los derechos de los autores, era realmente un instrumento diseñado para permitirle a la corona inglesa regular la incipiente industria de la imprenta, y con ello ejercer un control de censura sobre esta poderosa y revolucionaria tecnología que permitía difundir ideas masivamente, lo cual resultaba peligroso para la monarquía[5]. Resulta fascinante cómo este tema es tan actual hoy, en el que observamos importantes tensiones entre la libertad de pensamiento, los neuroderechos, la expresión libre, la moderación de contenidos, la censura y aquello que Madison denominó la *captura de la democracia por una facción egoísta*[6], que bien ha retratado el Doctor Pedro Salazar cuando se ha referido a la kakistocracia contemporánea que detenta simultáneamente el poder político,

3 M. Witty, *Athenaeus describes the most ancient intellectual property, Prometheus*, Pluto Journals, Leicester, Inglaterra, 2018.

4 Este es el término usado en México por el Doctor David Rangel Medina, y es el que prefiero frente al de "propiedad intelectual", porque refleja la verdadera naturaleza de esta disciplina jurídica, cuyo primer principio y fin último es social. Rangel Medina, David, *Derecho Intelectual*, Universidad Nacional Autónoma de México, México, 1998.

5 Probablemente no sea una coincidencia que la monarquía inglesa haya sobrevivido mientras que la corona francesa, que no logró imponer una legislación de censura tan poderosa, sucumbió ante la proliferación de las ideas ilustradas. Sería hasta el advenimiento de la bisnieta de la imprenta, es decir, la Internet, que la humanidad comprendiera la verdadera dimensión de las monarquías alrededor del mundo, en su carácter de objeto decorativo, en lugar de su otrora pretendida naturaleza divina.

6 Hamilton, Alexander, *et al*, *The Federalist Papers*. Signet Classics, an Imprint of New American Library, a Division of Penguin Group, Estados Unidos de América, 2005.

económico e ideológico: "La mancuerna Trump-Musk y la pandilla que los rodea constituyen el mayor peligro para nuestras libertades que haya existido en la historia moderna de la humanidad. La amalgama de poderes que detentan, atada con su deriva fascistoide, nos deja a la intemperie y sin defensas"[7].

Independientemente de lo anterior, Halbert introduce una noción que retomaremos más adelante, en el sentido de que "las batallas en el derecho de autor reflejan las intenciones mutuamente excluyentes de mantener la autoridad y el control y la necesidad de relajar el control para facilitar la circulación (de una obra). La postura de cada persona en este debate depende de su incentivo económico para compartir o bien, controlar información". Además, retoma la noción de Brad Sherman y Alain Stowel referente a que "el derecho de autor es un cuerpo normativo que debe ponerse 'en contexto' lo que significa que actualmente está siendo sometido al escrutinio histórico al que ya han sido sometidos otros aspectos del Derecho", resaltando la importancia de considerar la realidad jurídica de un momento y lugar determinado en la historia para entender el sistema normativo propuesto por el Derecho para regularla[8].

Ahora bien, las creaciones de la mente, en específico, las obras artísticas (sin importar su naturaleza), como el propio Derecho, han sido objeto de una transformación impulsada por el desarrollo tecnológico y las herramientas o medios puestos a disposición de los creadores. Randall Packer y Ken Jordan refieren a la relación entre artistas y científicos, ya que los primeros, al igual que los juristas, se allegan de los avances de los últimos para implementar mejoras en sus procesos y resultados: "las generaciones posteriores de artistas buscaron, y encontraron, formas integradas y estrategias interdisciplinarias para expresar su preocupación por la conciencia individual y social y los estados extremos de la experiencia subjetiva. En los años transcurridos desde la Segunda Guerra Mundial, los científicos han perseguido la informática personal y la interactividad persona-ordenador como vehículos para transformar la conciencia, ampliar la memoria, incrementar el conocimiento, amplificar el intelecto y potenciar la creatividad. Las aspiraciones idealistas e ideológicas de ambos grupos han dado lugar a un nuevo medio que hace hincapié en la elección individual, la libre asociación y la expresión personal". En específico, Packer y Jordan mencionan como adelantos relevantes para la conformación de la multimedia al Internet, el

7 Salazar, Pedro *Kakistocracia 'reloaded'*, columna en *El Financiero*, 29 de enero de 2025, consultable en https://www.elfinanciero.com.mx/opinion/pedro-salazar/2025/01/29/kakistocracia-reloaded/.

8 D. J. Halbert, *Intellectual Property in the Information Age*, Quorum Books, Connecticut, Estados Unidos, 1999.

CD-ROM, la cámara digital y la realidad virtual, entendiéndose a la multimedia como la forma de arte concatenado, al que se refirieron artistas como Wassily Kandinsky como el encorsetamiento de artes que al ser usadas propiamente darán lugar al arte verdaderamente monumental[9].

El Derecho Intelectual ha sabido adaptarse a los cambiantes estándares que impone la nueva realidad jurídica reformada por la tecnología. Si bien, como hemos comentado anteriormente, en el mundo occidental moderno el Derecho de Autor surgió, sin así planearse expresamente, a partir de un mecanismo de control editorial sobre las imprentas, lo cierto es que ésta, la más noble de las disciplinas jurídicas, ha tenido una impresionante capacidad evolutiva de adaptación a la realidad que la ha rodeado históricamente.

Como respuesta a las revoluciones tecnológicas de nuestra civilización, el Derecho ha tenido que reaccionar, algunas veces con ecuanimidad y otras convulsamente, al advenimiento de nuevas tecnologías. A veces ha tenido que encontrar soluciones forzadas debido a que, si bien el Derecho prevé supuestos fácticos y les otorga consecuencias jurídicas, lo cierto también es que muchas controversias no están previstas en el supuesto normativo y por lo tanto el corpus normativo no establece una consecuencia jurídica aplicable a esa situación particular. Este tipo de controversias son usualmente resueltas a través de la jurisprudencia, en aquellos casos en que los juzgadores se aventuran a asumirse como tales y hacen un esfuerzo interpretador e integrador que permita solucionar problemas nuevos con las herramientas jurídicas existentes.

No creo que sea necesario crear un Derecho de la Inteligencia Artificial (IA), de la misma manera que el Juez Frank Easterbrook ha dicho que no se necesita un Derecho del Caballo,[10] cuando se refería a la moda del supuesto Derecho del ciberespacio[11] La IA está permeando, transversalmente, un sinnúmero de aspectos de nuestra civilización como hoy la conocemos. Considero que una buena parte de las cuestiones jurídicas que se suscitan alrededor de la

9 R. Packer y K. Jordan, *Multimedia From Wagner to Virtual Reality*, Norton & Company, Nueva York, Estados Unidos, 2001.

10 Easterbrook, Frank H., *Cyberspace and the Law of the Horse*, University of Chicago Legal Forum: Vol. 1996, Artículo 7, consultable en https://chicagounbound.uchicago.edu/uclf/vol1996/iss1/7/.

11 Recientemente en México, hemos visto una proliferación de iniciativas de ley que definen "ciberespacio" como un término jurídico, con naturalezas muy diferentes y hasta contradictorias, siendo que el término proviene de la novela de ciencia ficción *Neuromancer*, de Gibson: Gibson, William, *Neuromancer*, Ace Science Fiction Books, Nueva York, EUA, *1984*.

Inteligencia Artificial Generativa (IAG) pueden resolverse con las herramientas jurídicas existentes. Insisto, los jueces tendrán que querer asumirse como juzgadores y no como autómatas del Derecho. En algunos ámbitos, como el penal o el correspondiente al Derecho Administrativo Sancionador[12], sí que se necesitará normatividad nueva, por tratarse de entornos de aplicación estricta, en donde no se permite la interpretación, por ejemplo en el caso de los delitos o las infracciones administrativas.

Por supuesto, en el mediano y largo plazo, serán bienvenidos los desarrollos legislativos, de política pública y de gobernanza que giren alrededor de la IA. Similarmente a lo expresado por el Juez Easterbrook, el Doctor Cossío ha utilizado la analogía del Derecho del Ferrocarril[13]: cuando se creó este nuevo medio de transporte, no existían leyes ni precedentes judiciales que regularan los hechos jurídicos e ilícitos que surgieron como consecuencia de su adopción. Ello no resultó en la creación de un régimen jurídico del tren. Simplemente los jueces comenzaron a resolver las nuevas controversias suscitadas con motivo de esta innovación tecnológica. Con el tiempo, se conformaron criterios jurisprudenciales y se emitieron leyes particularmente dirigidos al entorno ferrocarrilero (por ejemplo, en los ámbitos laboral, penal, contractual, regulatorio).

También, espero que algún día podamos ponernos de acuerdo en el contexto internacional sobre una regulación y gobernanza universales, uniformes de la IA.

Sin embargo, el panorama geopolítico parece no ser halagüeño. La polarización entre las perspectivas de las potencias mundiales tiene sus raíces entre la competencia industrial que se avecina en materia de IA. Si hubiera un consenso entre los criterios de las diferentes potencias tecnológicas, existiría una homogenización que crearía *lex mercatoria* que permitiría definir un "campo de juego" en igualdad de condiciones para los agentes más relevantes en la industria de la IA, incluyendo las autoridades regulatorias internacionales y nacionales.

Por ejemplo, la Unión Europea ha dejado claro con la entrada en vigor de diversos lineamientos y ordenamientos legales (cuyo epítome es la Ley de IA publicada en el 2024) que su enfoque en el desarrollo y uso de IA debe ser

12 Ver jurisprudencia 2a./J. 124/2018 (10a.), NORMAS DE DERECHO ADMINISTRATIVO. PARA QUE LES RESULTEN APLICABLES LOS PRINCIPIOS QUE RIGEN AL DERECHO PENAL, ES NECESARIO QUE TENGAN LA CUALIDAD DE PERTENECER AL DERECHO ADMINISTRATIVO SANCIONADOR, disponible en https://bj.scjn.gob.mx/doc/tesis/4fhyMHYBN_4klb4HgrC4/%22Derecho%20administrativo%22.

13 Cossío Díaz, José Ramón y Tsuru, Kiyoshi, *Mesa redonda: ¿Quién es responsable por actos cometidos con Inteligencia Artificial?*, Barra Mexicana, Colegio de Abogados, Ciudad de México, México, Septiembre 2023.

antropocéntrico, teniendo como inicio y fin el respeto a los derechos humanos, estableciendo principios rectores y remarcando posibles riesgos y sesgos que pueden derivarse del uso y/o desarrollo no regulado de Inteligencia Artificial.

Otros países como China o Estados Unidos también han fincado de manera certera su postura al respecto, coincidiendo paradójicamente en que al momento no es necesario sobre-regular la IA, para dejar así paso libre a la innovación en la materia.

Mustafa Suleyman se refiere a ello en su capítulo "Orgullo nacional, necesidad estratégica" en el que describe la importancia que tiene para China su agenda del 2030, misma que apunta a que el gigante asiático sea el referente global en todo lo relacionado a la IA[14]. La anterior visión contrasta directamente por la visión norteamericana, que busca esencialmente lo mismo, ser punta de lanza en la materia. Esta nueva guerra fría quedó en evidencia tras la liberación del modelo chino "DeepSeek" cuyo estreno global resultó en una pérdida insólita para la Bolsa de Valores de Nueva York de $593 billones de dólares del valor de Nvidia, fabricante de chips con sede en Santa Clara, California[15], misma que ante la desesperación del gobierno americano (con todo y sus últimos anuncios respecto a su política en torno a la Inteligencia Artificial), fue seguida de restricciones de uso del modelo DeepSeek en diversos sectores (tanto públicos[16] como privados[17]) de Estados Unidos.

Examinemos un ejemplo que nos proporciona el Derecho Comparado: la sentencia de la Suprema Corte de Justicia de los Estados Unidos de América en el caso Burrow-Giles Lithographic Company contra Napoleón Sarony en 1884. Como contexto de aquella época, el daguerrotipo, como se le denominó al aparato y procedimiento de fotografía de Louis Daguerre, había sido inventado y presentado aproximadamente 45 años antes de la emisión de la sentencia referida. El daguerrotipo, como herramienta de creación artística enfrentó toda clase de desaprobación y críticas, en específico de los pintores, quienes concebían al daguerrotipo como una amenaza al arte "verdadero" del lienzo y el pincel. ¿Cómo podría ser posible que el mero acto de oprimir un botón constituyera un proceso creativo, comparable al trabajo de dibujar bocetos o

14 M Suleyman. "The Coming Wave" Norton & Company, Crown New York, Estados Unidos, 2023.

15 https://www.reuters.com/technology/chinas-deepseek-sets-off-ai-market-rout-2025-01-27/

16 https://www.cnbc.com/2025/01/28/us-navy-restricts-use-of-deepseek-ai-imperative-to-avoid-using.html

17 https://www.business-standard.com/technology/tech-news/deepseek-s-ai-restricted-by-hundreds-of-companies-amid-concerns-over-leak-125013100085_1.html

pintar óleos? Uno de sus críticos en específico, Paul Delaroche, pintor francés del siglo XIX, acuñó la frase "Desde hoy, ¡La pintura está muerta!" al ver el primer daguerrotipo. Una cosa es cierta: la fotografía ciertamente no extinguió a la pintura, pero de alguna forma sí transformó la manera en la que la sociedad "mira, comprende y explora el mundo"[18].

La relevancia del caso para el análisis autoral de la IAG es bastante obvia: el quid de la cuestión es si puede considerarse que el producto resultante de una operación realizada con herramientas de IAG constituya una obra, y el accionante de dicha herramienta un autor. El caso se ha citado y discutido en foros y aulas alrededor del mundo, tanto por entusiastas como por detractores de la Inteligencia Artificial.

Los hechos del caso consisten en que el fotógrafo Napoleón Sarony tomó un retrato del escritor Oscar Wilde, misma que tituló "Oscar Wilde No. 18" en la que aparecía el escritor sentado, con su traje y corbata, y cubierto con un abrigo, su brazo izquierdo recargado sobre su pierna, sosteniendo su mejilla, con su cara ligeramente inclinada. La empresa litográfica Burrow-Giles Lithographic reprodujo la fotografía de Sarony sin su autorización y la utilizó con fines comerciales, ante lo cual, este último demandó a la empresa litográfica reclamando infracción de derechos de autor. En ese momento, la fotografía no estaba consideraba como obra autoral en la legislación estadounidense.

El ministro Miller sentenció el caso, estableciendo que las fotografías son consideradas obras en términos de la legislación autoral, ya que éstas no se limitan a los textos escritos, y especialmente debido a la actividad humana que subyace a la toma de la fotografía, en específico, la postura y vestimenta del modelo retratado, la iluminación, la escenografía, los textiles empleados en el retrato, y en general, la composición de la fotografía misma, elementos que constituyen decisiones creativas del autor que se traducen en originalidad.

18 C. Bellinetti, "Today, Painting is Dead: Photography's Revolutionary Effect", "Art & Object" Carolina del Norte, Estados Unidos, 2019.

Es entonces evidente que estamos ante un momento en la historia (tanto desde la perspectiva jurídica, como la perspectiva tecnológica-cultural) en la que se avecinan cambios de paradigmas. Por lo referido anteriormente, recalcando la rápida velocidad de evolución de la actual tecnología, y por otro lado, el miedo o la aprehensión del sector jurídico de regular algo que todavía no entiende en su totalidad, y que sabe que sin importar su decisión, se está anticipando al desarrollo futuro inmediato de estas tecnologías y sus posibles nuevos alcances o riesgos, los legisladores alrededor del mundo han tomado una noción de urgencia en regular la inteligencia artificial.

Aclarando el paradigma en el que nos encontramos, autores como Ian Ayres y Jack Balkin[19] han intentado subsumir el reto de regular la inteligencia artificial haciendo uso de otras regulaciones existentes en la mayoría de las jurisdicciones, tales como el deber de los mandatarios o apoderados y el deber fiduciario, bajo el entendido que las herramientas, programas o aplicaciones de inteligencia artificial no pueden ser consideradas como imputados o actores responsables de cierta conducta jurídica, sino que son sus creadores y usuarios quienes deben ser responsables por el uso que realicen de éstas. Para lo anterior proponen establecer las siguientes dos nociones legales frente a la regulación de inteligencia artificial, "la primera estrategia es que la ley a veces atribuye intenciones a una entidad. Por ejemplo, en el caso de un acto ilícito intencional, la ley atribuye intenciones a los actores. Se presume que los actores que lesionan a otros han tenido la intención de que sus acciones tengan consecuencias. La segunda estrategia obliga a los actores a cumplir una norma de comportamiento, normalmente de razonabilidad, independientemente de sus intenciones reales. Esta norma de comportamiento es externa a las intenciones reales de los actores." (por ejemplo, el debido cuidado). A través de estas dos perspectivas se regulan, primero, las conductas que sin importar la intención de su actor, la ley regula esta conducta y le impone una sanción y segundo, las conductas que sin importar la intención de su actor, la ley cataloga la conducta con base en un criterio objetivo (con base en las circunstancias hipotéticas alrededor de su actor) y las regula y sanciona en caso de no adecuarse a dicho criterio.

El punto principal detrás del trabajo de Ayres y Balkin es destacar cómo el Derecho se ha enfrentado anteriormente con este tipo de retos y cómo existen ya desarrolladas en la mayoría de las jurisdicciones ciertos mecanismos y conceptos jurídicos que podrían servir para entender el fenómeno

19 I. Ayres y J. M. Balkin, "The Law of AI is the Law of Risky Agents without Intentions", 2024.

tecnológico detrás de la inteligencia artificial y la forma en la que éste colisiona con la regulación.

II. LA INTELIGENCIA ARTIFICIAL

El término "inteligencia artificial" se acuñó desde la conferencia sobre informática del Darmouth College de 1956, cuando se propuso una clínica para estudiar este fenómeno, entendiéndolo como aquél relacionado con actividades humanas que pueden ser suficientemente específicas para ser simuladas por una máquina[20]. En principio, debemos desmitificar a la Inteligencia Artificial (IA) para dimensionarla en su contexto justo, despejando sus características y cualidades, con objeto de efectivamente conocer los alcances de esta herramienta tecnológica. La IA ha sido definida como "un campo de la ciencia y la ingeniería que se ocupa de la comprensión, desde el punto de vista informático, de lo que se denomina comúnmente como comportamiento inteligente. También se ocupa de la creación de artefactos que exhiben este comportamiento"[21].

Más de 60 años desde que el campo de estudio computacional conocido como inteligencia artificial fuera concebido, esta tecnología está cambiando radicalmente nuestra civilización a un punto tal en que naciones enteras consideren dicha tecnología como potencialmente riesgosa o sujeta a regulación y vigilancia.

En la actualidad los diversos cambios tecnológicos han dejado de versar respecto de medios de difusión y reproducción, pues desde la invención del ordenador y la tecnología informática, en específico la inteligencia artificial los desarrollos se han centrado específicamente en la colaboración para la creación de materiales diversos, e incluso para su creación y automatización, situación que ha generado diversos cuestionamientos relacionados con los Derechos de Autor, de los cuales abordaremos algunos en este documento.

Como hemos referido anteriormente, los contextos tecnológicos han generado vertiginosos cambios respecto de la difusión, distribución, explotación e incluso producción de diversas obras.

No obstante, los cambios tecnológicos y su repercusión en la explotación y protección de obras y materiales no necesariamente es algo novedoso para los

20 Dartmouth College, "Artificial Intelligence (AI) Coined at Dartmouth" https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth

21 R. Pino Diez., R. Gómez Gómez., N. Abajo Martínez., *Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva*, Universidad de Oviedo, Servicio de Publicaciones, Oviedo España, 2001.

derechos de autor y conexos, pues como hemos ilustrado a través del ejemplo del Estatuto de la Reina Ana de 1710 y la sentencia de Burrow-Giles Lithographic Company contra Napoleon Sarony de 1884.

Otro ejemplo de ello es la búsqueda de marcos regulatorios internacionales que promovieran el respeto a los derechos de los creadores, dando lugar a la unión de Berna[22] por virtud del desarrollo y modernidad en las vías de comunicación y el transporte de obras en el mundo.

Por otro lado, han existido aquellos elementos tecnológicos que han revolucionado la manera de realizar negocios, ejemplo de ello es la creación de máquinas mecánicas que reproducían música, ante la cual los autores hicieron reclamos por el uso de sus obras, una situación que no se vio reflejada hasta el Acta de Berlín de 1908 del Convenio de Berna para la Protección de Obras Artísticas y Literarias, situación que derivó en la creación de figuras como las licencias obligatorias[23] o derechos de simple remuneración para autores, intérpretes y ejecutantes.

Debido a que la funcionalidad de la IA se encuentra relacionada con las actividades intelectuales, es que esta tecnología es potencialmente aplicable en una infinidad de ámbitos de actividad humana. Desde los años 80 han existido aplicaciones comerciales de la IA en áreas como la producción, control de procesos, y contabilidad;[24] no obstante, la inteligencia artificial generativa tomó auge en las actividades humanas relacionadas con la creatividad desde 2014.

Diversos avances tecnológicos y computacionales debieron ocurrir a fin de que pudiesen generarse materiales asimilables a obras por medio de computadoras, de los cuales se hablará breve y superficialmente en los párrafos siguientes.

---

22 Reunión de países interesados en consensuar un marco regulatorio que protegiera la creación autoral de obras originales, misma que culminó en la redacción del Convenio de Berna de 1886, tratado internacional actualmente firmado por 181 países.

23 Definidas por la Organización Mundial de la Propiedad Industrial (OMPI) como "Limitación sobre los derechos de los titulares de derechos de autor o derechos conexos, cuya esencia es que no depende del titular de un derecho particular si se concede o no autorización para llevar a cabo actos protegidos por el derecho en cuestión. (...) Deben concederse de forma obligatoria por parte de los titulares de los derechos o, en su lugar, por una organización de gestión colectiva o por la autoridad competente. Se diferencia de una licencia legal en que en esta última el usuario ni siquiera tiene que solicitar una licencia, ya que la legislación autoriza directamente el uso en cuestión y solamente debe pagarse una remuneración." Existen ordenamientos jurídicos tales como el Convenio de Berna y en la Convención Universal sobre el Derecho de Autor, en los que se prevén licencias obligatorias.

24 S. Navas Navarro, *Inteligencia artificial, tecnología y derecho*, Tirant lo Blanch, España, 2017.

Los ochentas trajeron un movimiento denominado conexionismo, movimiento de la cibernética que abordó los problemas desde un punto de vista cognitivo, pues se aplicó el modelo de redes neuronales, permitiendo la depuración de aprendizaje profundo mediante técnicas de representación distribuida y algoritmos de retro propagación, situación que ha sido descrita más como una retroalimentación, lo cual se ha referido como una retroalimentación sofisticada, para dar paso en 2006 a que el investigador Geoffrey Hinton desarrollara el "*deep learning*".[25]

Si bien el *deep learning* fue utilizado principalmente para reconocimiento de imágenes, esta tecnología daría paso para que Ian Goodfellow desarrollara lo que ahora es denominado como redes generativas adversarias[26], las cuales permiten la generación de imágenes mediante un sistema que utiliza dos redes de competencia, una discriminadora y otra generadora. El sistema funciona cuando la red generadora constantemente crea imágenes, y la red discriminadora analiza si dichas imágenes son "falsas" o son parte de información provista al sistema o no, permitiendo que con base en los resultados ambas redes mejoren al punto de crear diversas imágenes con similitud a la realidad, no obstante ser falsas.

Lo anterior permitió que los desarrollos computacionales llegaran al punto de crear materiales suficientemente buenos como para pasar como obras realizadas por personas, incluso para imitar fotografías de paisajes o personas creadas artificialmente.

Desde entonces se ha planteado la pregunta, ¿los materiales creados por medio de herramientas como la inteligencia artificial generativa generan derechos de autor? De ser así ¿Quién es el autor de dichos materiales?

Al respecto, y previo a entrar en materia de análisis del panorama internacional, dos autores de suma importancia para los derechos de autor han descrito lo que debe contener una obra para ser objeto de protección. Por un lado, Isidro Satanowsky refería que "una obra intelectual es una construcción de la mente, de carácter completo, integral y unitario. Debe tener, asimismo, una existencia propia, ser completa por sí misma y presentar al espíritu un sentido, una significación a la que nada haya que agregar. Resulta menester que posea

25 B. Caldas Vianna "Generative Art: Between the Nodes of Neuron Networks" AI, Arts & Design: Questioning Learning Machines. Número. 26.

26 I. Goodfellow, "Deep Learning", The MIT Press, Massachusetts, Estados Unidos, 2016.

una individualidad y requiere no sólo una tarea activa y decisiva, sino una labor creadora y autónoma, que dé forma y vida espiritual a una idea"[27]

Por otro lado, Delia Lipszyc se centra en el concepto de individualidad como requisito de existencia de una obra, "en materia de derecho de autor, la originalidad reside en la expresión o forma representativa-creativa e individualizada de la obra, por mínimas que sean esa creación y esa individualidad" algunos autores prefieren utilizar el término individualidad, en lugar de originalidad, por considerar que expresa más adecuadamente la condición que el derecho impone para que la obra goce de protección: tenga algo de individual y propio de su autor[28].

III. PANORAMA INTERNACIONAL

Empezando por los Estados Unidos de Norteamérica, su Oficina de Derechos de Autor (USCO por sus siglas en inglés) ha sostenido en múltiples ocasiones que, el material generado con intervención absoluta de la inteligencia artificial carece del requerimiento esencial[29] para ser registrable como una obra protegida por el derecho de autor.

Primero, en el caso[30] de "*A Recent Entrance to Paradise*", obra de dibujo creada por el sistema de IA Creativity Machine, a su vez creado por el controvertido Stephen Thaler, el Órgano Revisor de la USCO determinó en febrero de 2022 que la solicitud de registro no contenía evidencia de intervención suficiente por un ser humano para la realización de la obra, razón por la cual la USCO determinó que era imposible inscribir una obra que no había sido realizada por un autor persona física.

Derivado del caso citado anteriormente, la USCO refirió que las obras protegibles bajo su jurisdicción son aquellas que representan el fruto de la labor intelectual y que se encuentran basadas en el poder creativo de la mente humana. En esta misma respuesta, la USCO refirió que, consecuentemente, los productos de un simple proceso mecánico, sin ninguna intervención de un autor humano, no son protegibles en principio.

---

27 (conf. Isidro Satanowsky, «Derecho Intelectual», Ed. Tipográfica Editora Argentina (TEA), Bs. As.1954, p. 154)

28 LIPSZYC, Delia...Derecho de Autor y Derecho Conexos. UNESCO/CERLALC/ZAVALIA. Buenos Aires. 1993. Pág. 65.

29 Según se establece en el "'Capítulo 306: Requisito de Autoría Humana' previsto en el 'Compendio de prácticas de la Oficina de Derechos de Autor de los Estados Unidos', Tercera Edición".

30 Thaler v. Perlmutter, No. 22-1564 (D.C. Circ. 2023)

La USCO decidió, en otro intento de registro de obra de febrero de 2023, en relación con la negativa para proteger ciertas imágenes generadas por el servicio de inteligencia artificial "Midjourney" en el caso del cómic titulado "*Zarya of the Dawn*", en el que la USCO decidió que el proceso que herramientas como la referida utilizan, no supone creatividad humana alguna, toda vez que si bien los comandos ingresados por el usuario humano al sistema de inteligencia artificial (también conocido como '*prompt*') pueden influenciar el resultado que la inteligencia artificial proveerá, no es posible conocer el resultado de esta por virtud de la creatividad humana, es decir, no interviene la creatividad humana sino un sistema de aprendizaje profundo en conjunto con un sistema adversarial de inteligencia artificial que aleatoriamente arroja resultados, (a diferencia de la utilización de herramientas como una cámara fotográfica, en donde el autor tiene control creativo "Chispa Creativa"[31] generadora de un resultado), el uso de inteligencia artificial elimina cualquier elemento de creatividad humana requerida para la protección jurídico autoral.

La USCO, a través de su criterio y lineamientos que ha publicado, ha sostenido que si bien una obra creada por IA no puede ser registrable en términos de la legislación aplicable, lo que sí es protegible son aquellas porciones de una obra en las que haya intervenido la impronta del autor, sin importar que estas porciones se compensen de contenido creado por IA para componer la obra, por ejemplo, en *Zarya of the Dawn*, se decidió que si bien las ilustraciones del cómic fueron generadas por IA y por ende, no son protegibles, otros aspectos como los diálogos, el acomodo de las imágenes, entre otros, sí son fruto de la impronta del autor, por lo que esas porciones del cómic si resultan protegibles por el derecho de autor.

De esta manera, la USCO establece un criterio gradual por medio del cual requiere que los solicitantes de un registro de derechos de autor, declaren qué partes de su obra (si es que aplica) fueron creadas haciendo uso de herramientas de IA, para de esta manera no dejar sin protección elementos de estas obras que cumplen con los requisitos de ley para ser considerados obras, y en consecuencia, estar protegidos por la ley autoral.

La USCO, además de los casos anteriormente citados ha emprendido una iniciativa que data del año 2023 por medio de la cual se han recibido comentarios de todo el mundo sobre Inteligencia Artificial. Como parte de esta iniciativa, la USCO se propuso publicar un reporte de tres partes, la primera se especializa en el tema de "Réplicas Digitales" o *deepfakes*, la segunda parte

31 Feist Publications, Inc. v. Rural Telephone Service Co., 499 U.S. 340 (1991)

aborda la registrabilidad de obras generadas por IA y la tercera parte abordará las implicaciones legales en el entrenamiento de modelos de IA con obras protegidas por el derecho de autor[32].

En su segunda parte del Reporte, la USCO comparte ciertos criterios, por ejemplo, que la ley actual de derechos de autor americana es suficiente, apta y no requiere ninguna modificación para regular las obras generadas por (o haciendo uso de) Inteligencia Artificial. Además, desarrolla con mayor profundidad el sistema de graduación que ya esbozaban decisiones de la USCO citadas con anterioridad en el presente, en específico, aseverando que la protección autoral requiere forzosamente impronta del autor, sin embargo, también reconoce que la IA puede ser usada como una herramienta (algo similar a lo decidido en cuanto al uso de la cámara fotográfica en el caso Burrow-Giles Lithographic Company contra Napoleón Sarony) que interactúa con un humano, siguiendo las instrucciones del usuario detalladas en los *prompts* introducidos para conseguir el resultado deseado.

La USCO sin embargo deja muy en claro que los análisis de registrabilidad deben ser analizados y decididos caso por caso. La USCO subraya las diferencias en su generación y distingue entre las obras referidas anteriormente, conceptualizándolas como "obras asistidas por IA", separándolas de aquellas obras que sean generadas por IA sin ninguna interacción humana, dejando fuera de la ecuación la impronta de un autor humano en su creación, y por ende, descalificándolas para su registrabilidad y definiéndolas como "obras generadas por IA".

El criterio de registrabilidad de la USCO entonces queda aún más definido con su segundo reporte, estableciendo claramente que las "obras generadas por IA" no tienen detrás decisiones o controles creativos, que típicamente conforman los procesos de creación de cualquier tipo de obras. Estos procesos son realizados por algoritmos computacionales que buscan responder al *prompt* con base en la información con la que fueron alimentados. Es evidente que en este proceso no existe ningún momento en el que se pudiera argumentar la impronta de un autor, por lo que no hay una chispa creativa, ya que el algoritmo asignará un determinado resultado al *prompt* recibido,

32 https://www.copyright.gov/ai/

sin intervención de ninguna persona. Para ilustrar lo anterior, la USCO usó la siguiente imagen en su reporte:

Tomando en cuenta este criterio, la USCO ya ha aprobado el registro de una obra generada por IA, en específico se trata de la obra "*A Single Piece of American Cheese*" de Invoke, una plataforma de Inteligencia Artificial Generativa. Invoke ha referido que en su solicitud de registro a la USCO remarcó que la creación de su obra fue lograda a través de un proceso de varias etapas que conllevó un perfeccionamiento iterativo y la toma de decisiones creativas por parte de personas, lo anterior para solicitar la protección de la selección, refinamiento y arreglo de la disposición del material generado por Inteligencia Artificial en el producto final. Como nota, esta obra ya había sido sometida a la revisión de la USCO en septiembre de 2024, revisión a la cual le recayó una negativa, misma que fue revocada en enero de 2025, poco después de la publicación del segundo reporte de la USCO.

Como conclusión del panorama estadounidense cabe destacarse que actualmente existen más de veinte diferentes juicios cuyas sentencias probablemente terminen de definir el criterio que seguirán cortes y autoridades de aquel país por igual en lo referente a material generado por Inteligencia Artificial.

Cambiando de jurisdicción y concentrándonos en un territorio igualmente importante en la carrera universal de desarrollo de tecnologías, en China se ha construido un sistema jurídico que obedece a una política industrial que busca proteger la inversión económica de quienes crean nuevas tecnologías.

La óptica china se ve reflejada en tres diferentes decisiones jurisdiccionales de Cortes de Shenzen, Guangzhou y Beijing.

En el primer caso, Shenzen Tencent Co., Ltd., contra Shanghai Yinxun Technology Co., Ltd, decidido en 2020 por la Corte Civil del distrito de Nanshan en Shenzen, falló a favor de Tencent, defendiendo que si bien los reportes que su herramienta de IA llamada

Dreamwriter generaba al terminar las labores de la bolsa de valores, el personal de Tencent acomodaba y gestionaba la información contenida en estos reportes, lo que permitía que estos fueran considerados como obra por tal motivo y en consecuencia se sancionó su reproducción no autorizada realizada por el demandado.

El segundo caso, Li contra Liu, decidido por la Corte de Internet de Beijing, estableció el criterio de que las obras generadas por IA que sean producto de *propmts* especializados (compuestos por limitantes, prohibiciones, lineamientos, directrices creativas, entre otros elementos) son protegibles en función de la impronta que el autor determinó en dicho *prompt*. En este caso, se decidió sobre si ilustraciones de mujeres que fueron producidas a través *Stable Difussion* usando *prompts* de casi 300 palabras eran sujetas a protección autoral, ya que el demandante reclamó de la demandada el uso no autorizado de sus obras generadas por *Stable Difussion*.

En el tercer y último caso Shanghai Character License Administrative Co., Ltd contra la herramienta de IA llamada "Tab", la Corte de Internet de Guangzhou decidió que el operador de Tab, que pone a disposición un modelo de IA Generativa a través de su página web infringió el derecho exclusivo de Shanghai Character License Administrative Co. de reproducir al personaje "Ultraman" (licenciado por Tsuburaya Productions Co, titular primigenio de los derechos del personaje). En específico, porque al utilizar en el *prompt* la palabra Ultraman, el modelo de IA de Tab generaba una copia idéntica del personaje original.

Además, la Corte estimó que la plataforma también infringió en el derecho exclusivo licenciado para crear obras derivadas del personaje Ultraman, ya que a través de *prompts* cualquier persona podría generar variaciones del personaje haciendo uso de la herramienta de IA.

IV. PANORAMA EN MÉXICO

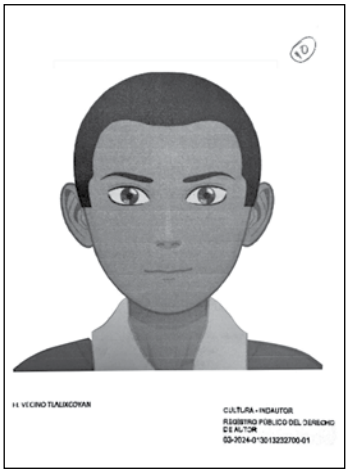
Habiendo analizado el panorama internacional, es menester analizar el panorama nacional.

En el ámbito legislativo, hay que mencionar que en la Legislatura anterior se presentaron más de 60 iniciativas de ley presentadas en las Cámaras de Senadores y de Diputados para regular, directa o indirectamente, algún aspecto de la IA, de las cuales, algunas contenían disposiciones en materia de propiedad intelectual (incluso, algunas fueron propuestas de reforma a la Ley Federal del Derecho de Autor). Ninguna de estas iniciativas ha sido fructífera, por lo que ninguna modificación en la materia ha sido implementada en nuestro marco jurídico nacional.

También debe enfatizarse que nuestra Constitución Política establece en su artículo 28 que "En los Estados Unidos Mexicanos quedan prohibidos los monopolios, las prácticas monopólicas, los estancos, las condonaciones de impuestos y las exenciones de impuestos en los términos y condiciones que fijan las leyes. El mismo tratamiento se dará a las prohibiciones a título de protección a la industria. (…) Tampoco constituyen monopolios los privilegios que por determinado tiempo se concedan a los autores y artistas para la producción de sus obras y los que para el uso exclusivo de sus inventos, se otorguen a los inventores y perfeccionadores de alguna mejora. (…)", lo anterior para destacar que el fin constitucional de la protección de la propiedad intelectual es incentivar la creación de obras, cuestión que es inaplicable a la Inteligencia Artificial ya que ésta no puede ser incentivada, y si se incentivara las empresas que desarrollen Inteligencia Artificial Generativa se estaría desvirtuando el fin último constitucional anteriormente referido.

El Tribunal Federal de Justicia Administrativa ha sentado jurisprudencia a través de la sentencia de su Sala Especializada en Materia de Propiedad Intelectual del Juicio Contencioso Administrativo Número 788/24-EPI-01-2, a través de la cual ha establecido su criterio, especificando que "aunque sea una persona física quien proporciona los lineamientos e instrucciones a la inteligencia artificial, la creación no deja de ser producto de la ejecución de los algoritmos que ésta emplea generándose un contenido artificial. Por ello, en términos de la Ley Federal del Derecho de Autor, no puede ser registrable una obra que no sea creada por una persona física, por el hecho de que se encuentra expresamente señalado de esa forma en ley porque solamente ésta es la persona capaz de crear una obra original (requisito exigido por ley) porque para tal situación es necesaria la creatividad humana." Este criterio permite que nuestro sistema pueda adaptarse a la graduación prevista por la USCO y abordada previamente, ya que si bien esta jurisprudencia resulta aplicable a contenido generado por IA, no se hace mención específica al caso en el que una obra esté compuesta tanto por material generado por IA, como por obras realizadas por personas (autores en ejercicio de su actividad creativa a través de su impronta).

Con base en el anterior criterio, se han rechazado los registros de obra solicitando protección de contenido generado por Inteligencia Artificial, como los siguientes:



V. PROPUESTAS

Es necesario que se establezcan criterios claros sobre cuándo es posible que un autor, persona física, pueda considerarse autor cuando crea una obra con la asistencia de una herramienta de Inteligencia Artificial Generativa. No estoy diciendo que un sistema informático debería considerarse como un autor. Claramente ello sería un sinsentido. Los reclamos de Thaler pretendiendo obtener protección autoral son un disparate.[33]

El Derecho Intelectual tiene un primer principio, fin último y vocación social. No constituye un derecho absoluto. No es un fin en sí mismo. Es un medio para llegar a un objetivo tendiente al bien común: incentivar a los creadores para que creen, mediante el otorgamiento de un privilegio temporal de explotación, de manera que se aumente el volumen del acervo cultural de la nación, pero, sobre todo, que se produzcan obras de mayor calidad, al permitirle al autor profesionalizarse y dedicarle todo su tiempo y atención a su obra. A los humanos se les puede incentivar. A las máquinas no.

Es claro que solo un ser humano puede ser autor. Ésta es la respuesta simple a la cuestión. El asunto que resulta más complicado es determinar qué grado de aportación del usuario es necesario para que pueda considerarse que la interacción de éste con la herramienta de IA constituye la creación original de una obra.

33 Ver Thaler v. Perlmutter, *supra*

Es innegable que los autores están usando, cada vez con mayor frecuencia (y con éxito), herramientas de IA. En la medida en que la tecnología vaya permitiéndole al usuario tener mayor control sobre la manera en que la IGA genera resultados, y que consecuentemente pueda haber mayor trazabilidad sobre la impronta del autor en la expresión particular de su obra, es decir, la relación inmediata y directa entre las decisiones creativas del usuario y la generación del contenido a partir de ellas.

Pareciera que no estamos aún en ese momento. Conforme al estado de la técnica, la generación de contenidos a partir de *promptings* del usuario no implica un control directo del usuario sobre la ejecución de la expresión del contenido en cuestión. El Derecho de Autor protege expresión, no idea. Por ahora, el proceso de transformación de una instrucción conceptual en un contenido (texto, imagen, música, video) con la ayuda de una herramienta de IGA sucede de manera automatizada, conforme al mecanismo de operación de Modelos de Lenguaje Extenso (LLMs).

Este análisis no debería tomarse como un dictamen inmutable. Si la tecnología evoluciona, de manera que permita al usuario tener control más directo y efectivo sobre la ejecución del proceso creativo, de manera que su impronta pueda imprimirse de manera clara e identificable sobre una obra, y que se garantice la trazabilidad de las decisiones creativas desde su concepción hasta la creación de la expresión, entonces podremos pasar al siguiente nivel de la discusión.

En un escenario ideal, los criterios sobre originalidad deberían estar uniformados transversalmente, por ejemplo, entre todos los países miembros del Convenio de Berna o del Acuerdo sobre los Derechos de Propiedad Intelectual Relacionados con el Comercio (ADPIC).

Otra alternativa podría ser la conformación de un nuevo régimen *sui generis* de protección a los contenidos hechos con la asistencia de herramientas de IAG. Nuestro sistema autoral ya incorpora, si bien de manera muy breve, un modelo *sui generis* de protección de bases de datos, similar al previsto actualmente para las bases de datos en el Capítulo IV del Título IV de la Ley Federal del Derecho de Autor, por medio del cual coexistan los derechos de autor de las obras protegibles creadas por personas y la protección sobre los contenidos sintéticos generados por IA. Este derecho tendría por objeto proteger no solamente los algoritmos y mecanismos computacionales que hacen posible la generación de contenido a través de estos sistemas, sino también a los datos utilizados en la creación de dichos modelos de Inteligencia Artificial. De esta manera, no se estaría otorgando la calidad de autor a un conjunto de algoritmos computacionales, ni de obra al material generado por éste, pero se estaría

protegiendo la inversión detrás de éstos, además de facilitar la identificación de materiales que podrían resultar infractores de obras protegidas bajo la normativa actual[34].

Este tipo de protección tampoco es ajena al derecho explorado. Basta con voltear a nuestro país vecino y sus precedentes jurisprudenciales la materia, específicamente al caso Feist Publications, Inc., contra Rural Telephone Services Co., resuelto por la Suprema Corte de los Estados Unidos. El contexto del caso es que Rural Telephone Services, una empresa que proveía servicios de telefonía compiló una guía con los nombres y los teléfonos de sus usuarios. Feist Publications copió parte de la información recopilada y publicada en las guías de Rural para crear sus propias guías. La Suprema Corte Estadounidense decidió que si bien las guías de Rural no cumplían con los requisitos establecidos en su ley para otorgarle la calidad de obra (y en consecuencia, estar protegida por los derechos de autor), el esfuerzo existente detrás de la selección, compilación y organización de información detrás de las guías de Rural merecía una protección ante la ley, esbozando en dicha resolución el concepto del sudor de la frente ("*sweat of the brow*") para conceptualizar el trabajo realizado por Rural y sus empleados que produjo las guías telefónicas en controversia.

Al contar con una protección *sui generis* como ésta, se estaría promoviendo la innovación, la protección a la inversión en la jurisdicción mexicana y se presentaría un incentivo para que creativos de todo el mundo utilicen esta jurisdicción como base para sus desarrollos tecnológicos y culturales.

En el terreno de las infracciones administrativas previstas por nuestro ordenamiento autoral, específicamente las llamadas infracciones en materia de comercio, y atendiendo a los principios del Derecho Administrativo Sancionador (por ejemplo, *nulla poena sine lege*), nos parece que debería de adicionarse al catálogo de infracciones en materia de comercio de forma específica, el uso o reproducción no autorizada para copiar o bien para crear obras derivadas a través de herramientas de IA de obras protegidas por el derecho de autor.

Por último y no menos importante, debe prestarse atención a la minería de datos y texto (conocida por sus siglas en inglés como TDM), misma que refiere al proceso computacional de extracción de información pública realizada por mecanismos computacionales (incluidos los *crawlers*, que automáticamente

34 Ver la presentación de mi análisis sobre un posible régimen de protección *sui géneris*, en el foro "Irrupción de Nuevas Tecnologías: ¿Cambios de paradigmas al Derecho de Autor?" convocado por el Instituto Interamericano del Derecho de Autor (IIDA) y celebrado el 7 de mayo de 2024 en las instalaciones de Ciudad Universitaria, de la Universidad Nacional Autónoma de México (UNAM), a instancias de la Maestra Carmen Arteaga.

buscan y extraen texto e información disponible en páginas web). Esta técnica también conocida como *scrapping* tiene por objeto nutrir las grandes bases de datos (*datasets*) con los cuales se entrenan los modelos de Inteligencia Artificial y de los cuales depende la calidad de los materiales que éstos puedan generar. Diversas jurisdicciones alrededor del mundo están evaluando la posibilidad de incluir los procesos de TDM en sus respectivas limitaciones al derecho de autor. Por ejemplo, en Estados Unidos se está argumentando (por parte de las compañías propietarias de modelos de IA en su gran mayoría) que los procesos de TDM deben de encuadrarse en la doctrina estadounidense del *fair use*. En nuestro país, esta excepción podría preverse en el artículo 148 de la Ley Federal del Derecho de Autor que prevé un listado de *numerus clausus* de limitaciones a los derechos patrimoniales.

Finalmente, nos parece de vital importancia subrayar la necesidad de empezar a construir un marco jurídico en la materia que esté construido con base en el análisis y evaluación de la registrabilidad de las obras caso por caso. Lo más importante radicará en definir el grado de participación de una persona que utilice Inteligencia Artificial como parte de su proceso creativo de manera que su impronta y originalidad sean implementadas en su obra.

CAPÍTULO SEIS

INTELIGENCIA ARTIFICIAL. UNA VISIÓN GLOBAL COMPARADA

LA INTELIGENCIA ARTIFICIAL EN EL MUNDO DIGITAL DEL SIGLO XXI. UNA VISIÓN EUROPEA Y LATINOAMERICANA

SUMARIO

I. Un preámbulo necesario: Revolución Digital y Digitalización **II.** La digitalización tiene un contexto geopolítico **III.** Los modelos de Digitalización **IV.** Sobrevivir al Tsunami regulatorio digital europeo: las reglas del juego **V.** Inteligencia Artificial: La próxima burbuja **VI.** El Apocalipsis que viene **VII.** ¿Están justificados estos temores? ¿que es y que no es la Inteligencia Artificial? **VIII.** Los riesgos reales de la Inteligencia Artificial y su regulación **IX.** La visión europea **X.** El modelo americano **XI.** Desafíos comunes **XII.** Desafíos éticos **XIII.** Un epílogo latinoamericano **XIV.** La Inteligencia Artificial

> CARLOS LÓPEZ BLANCO
Patrono de la Fundación Euroamérica. Presidente de la Comisión de Economía Digital de la International Chamber of Commerce

I. UN PREÁMBULO NECESARIO: REVOLUCIÓN DIGITAL Y DIGITALIZACIÓN

Empecemos con una pregunta, aparentemente trivial: oímos constantemente que nos encontramos ante una Revolución, la Revolución Digital, que, se dice, abre una nueva era solo comparable a que abrió la Revolución Industrial de finales del siglo XVIII.¿ es esto cierto ?

Son palabras mayores, y en un mundo como el actual en que las hipérboles son de uso común, cabe pararse a pensar si de verdad se dan los elementos necesarios para considerar que estamos ante una época tan trascendental como lo fue la revolución industrial y el surgimiento del capitalismo moderno.

Se dice y con razón que la irrupción de la tecnología, y sus consecuencias no es un fenómeno nuevo, sino que se remonta a cambios tan importantes, pero ya lejanos en el tiempo, como la irrupción de la informática, primera tecnología digital, o la explosión de la conectividad producida por la convergencia de la informática y las telecomunicaciones de la mano de la telefonía móvil.

No se puede negar la trascendencia que estos fenómenos han tenido pero cabe dudar que esto por si mismo nos permite defender la idea de que estamos ante una nueva revolución, la Digital, que va a producir en la economía y la sociedad efectos comparables a las que tuvo la Revolución Industrial que, no olvidemos, es origen sin el cual no podemos entender ni la economía ni el mundo moderno.

En mi opinión, solo podemos fundamentar esta afirmación si nos enfocamos en la que ha sido la segunda fase de la Revolución Digital, tras la conectividad, que no es otra que el nacimiento y desarrollo de la economía de los datos.

Se ha dicho –con el talento característico de los medios anglosajones para acuñar formulas brillantes– que los datos son el petróleo del siglo XXI y por

tanto una nueva materia prima para fundamentar una nueva economía de la misma manera que el crudo fundamento el desarrollo del capitalismo durante todo el siglo XX.

¿CUÁLES SON LAS CARACTERÍSTICAS DE ESTA SEGUNDA REVOLUCIÓN DIGITAL, LA ECONOMÍA DE LOS DATOS?

1. Hiperconectividad.
Estamos hablando de un momento posterior a la de la conectividad, fase anterior, donde a la conectividad universal de las personas se le suma la conectividad de las cosas, de todo, el IoT, fenómeno llamado a cambiar de manera radical buena parte de nuestra economía y sociedad.

2. Hiperglobalización.
De la misma manera esta economía favorece el desarrollo de una nueva globalización, que va más allá de la globalización comercial y se centra en el comercio (trafico) de datos que ya tiene un volumen mayor que el desarrollo las mercancías físicas. Cuando se habla del fin de la globalización se ignora que la verdadera base de la economía del futuro, la economía de los datos sigue siendo por definición un fenómeno global; más bien debería hablarse del fin de un modelo que dominó la economía mundial la segunda mitad del siglo XX.

3. Las barreras entre el mundo físico y el virtual están desapareciendo.
Este es un fenómeno de gran trascendencia, ya que hemos vivido hasta ahora con una economía dual: la de lo analógico y la de lo digital. Realidades distintas con reglas distintas. Hoy cada vez es más difícil distinguir las fronteras entre ambas y las empresas analógicas cada vez tienen más actividad digital y viceversa, dando lugar a una nueva realidad que no es hibrida sino un fenómeno claro de digitalización como nuevo paradigma.

4. La economía y la sociedad de los datos.
Como ya hemos dicho es un factor determinante al surgir un nuevo tipo de materia prima de la actividad económica que son los datos y, sobre todo, el valor que se puede extraer de ellos, ya que en esto sí los datos son como el petróleo: tener reservas de crudo no sirve para nada si no tienes capacidad de extracción y de refino. De la misma manera la dinámica de la economía de los datos, pronto la Economía sin adjetivos, se basa en la capacidad de las empresas de extraer valor de esos datos. Sin esto, ni hay algoritmos ni hay Inteligencia Artificial.

Si todos estos factores nos conducen a la innegable conclusión que estamos ante un cambio radical en nuestra economía y la actividad de las empresas no bastaría ello para decir que estamos ante una auténtica Revolución, que solo puede calificarse como tal si afecta no solo a la economía sino a todos los órdenes de la sociedad y la vida de las personas, incluyendo la política.

¿ES UNA AUTÉNTICA REVOLUCIÓN, COMPARABLE A LA INDUSTRIAL DE FINALES DEL XVII Y CON SUS MISMAS CONSECUENCIAS ECONÓMICAS, SOCIALES Y POLÍTICAS?

Todo ha cambiado en la economía, pero también en la sociedad.

Si comparamos nuestra actividad cotidiana de hace solo diez años con la actual veremos que todo ha cambiado en nuestra vida: como vendemos y compramos, como encargamos nuestra comida, como hablamos, como nos relacionamos, como leemos, como escribimos, como escuchamos música, como leemos o como pagamos. Todas las actividades de la vida cotidiana, con connotaciones económicas o no, han sufrido una profunda y a veces imperceptible transformación en los últimos años, de la mano fundamentalmente de las plataformas y esto nos permite afirmar que no estamos viviendo una revolución tecnológica, vivimos un cambio social.

Este es un aspecto fundamental, ya que si pretendemos establecer la comparación o la equivalencia con la Revolución Industrial en términos de importancia historia como comienzo de la sociedad capitalista y la sociedad digital respectivamente, debemos tener en cuenta que este tipo de transformaciones no se limitan a la economía sino que deben proyectarse a todos los órdenes de la vida social, transformando la manera en que la sociedad se desenvuelve y estas características se dan en la Revolución Digital.

Sentada la afirmación de que estamos no solo ante un cambio tecnológico o el surgimiento una nueva economía sino ante un auténtico cambio social no es ocioso entender en qué momento se encuentra ahora este proceso.

Debemos empezar formulándonos una pregunta básica ¿qué es la digitalización?

No es esta una pregunta trivial porque no hay una clara diferenciación entre dos nociones que son clave para entender toda la problemática que nos plantea la Revolución Digital que estamos viviendo: la diferencia entre Economía Digital y Digitalización, la primera entendida como aquella parte de la economía basa en la producción para el resto del sistema de tecnología, desde el software

hasta los servicios en la nube, de los servicios de consultoría tecnológica hasta el desarrollo de algoritmos. Frente a ello cuando hablamos de digitalización nos referimos al proceso de incorporación masiva de la tecnología en todos los sectores productivos y tamaños de empresa.

Economía digital y digitalización son vasos comunicantes y conceptos que tienen una relación simbiótica, pero es necesario distinguirlos y entender que no es en la economía digital sino en la digitalización donde nos encontramos no ante una mera disrupción tecnológica sino un gran cambio social. La llamada revolución digital no es un cambio tecnológico, es un cambio social.

También la diferencia entre Economía Digital y Digitalización marca la base de dos estratégicas de política económica (y de economía política) diferentes y, en algunos casos, compatibles: el fomento de la construcción de un sector tecnológico competitivo y el fomento del proceso de introducción de los medios digitales (proporcionados por la Economía Digital) en todos los sectores de la economía (y la sociedad). En el primer caso la estrategia de los Estados Unidos, en el segundo la de Europa.

Este es el momento en que estamos: hasta ahora, hemos vivido con dos economías: la analógica tradicional y la digital con reglas distintas que convivían en un mundo en transformación, pero todavía dual: por una lado la economía digital y por otro lado economía analógica, ambas con reglas distintas y dinámicas diferentes donde el problema principal será como establecer un régimen isonómico que impida que la nueva economía digital arrase la tradicional analógica.

ECONOMÍAS CON REGLAS Y DINÁMICAS DIFERENTES Y NO COMPITIENDO *FACE TO FACE.*

Pero ha sido el proceso de digitalización y su aceleración a partir de la pandemia el que ha roto la diferencia entre economía digital y digitalización,

El dique entre ambas se ha roto: la economía/sociedad digital está invadiendo la tradicional y eso es lo que conocemos como proceso de digitalización: cada vez más la adopción de la tecnología digital será un proceso creciente e indispensable para mantener la competitividad de las industrias tradicionales que tendrán que digitalizarse o morir. De manera que en un tiempo lo que hoy llamamos economía/sociedad digital será la única existente cuando se culmine el proceso de digitalización.

II. LA DIGITALIZACIÓN TIENE UN CONTEXTO GEOPOLÍTICO

A finales del siglo XIX vivimos *The Great Game*, la lucha de los imperios decimonónicos, las grandes potencias europeas, por el dominio colonial del Asia Central, la gran zona geoestratégica del momento e inspiradora de la magnífica narrativa de Kipling.

La segunda mitad del siglo XX, por su parte, apareció dominada por la guerra fría, donde la lucha de los imperios se centraba no en el dominio colonial sino en la supremacía militar.

La caída del muro de Berlín abrió una nueva situación que se preveía unipolar bajo la hegemonía americana y que ha devenido en un mundo más inestable, transitoriamente multipolar, pero en que aparece un nuevo contendiente por la hegemonía mundial, China.

HOY LOS ACTORES SON OTROS. ¿CUÁL ES EL CAMPO DE BATALLA?

The next great game, esta nueva lucha por la hegemonía geoestratégica no tiene por objeto ni la supremacía colonial del XIX ni la supremacía militar de la guerra fría. La supremacía tecnológica es el objeto la lucha por la hegemonía mundial en el siglo XXI y en este contexto, la Inteligencia Artificial (y la computación cuántica) no es solo la próxima frontera tecnológica de la Economía Digital y la digitalización: es el gran caballo de batalla por la hegemonía y el liderazgo mundial con consecuencias en el orden geoestratégico, político, económico y social.

Cabe preguntarse quién tiene la mayor ventaja para esta contienda y en ese contexto, EEUU tiene el liderazgo tecnológico y la capacidad de innovación, pero China tiene la capacidad de ejecución.

China ha hecho viable un modelo que creíamos imposible: un mundo digital cerrado y controlado pero al mismo tiempo capaz de competir e innovar, demostrando que la capacidad de ejecución es un factor crítico, más importante de lo que pensábamos y que permite a China rivalizar en términos de competitividad con una economía altamente innovadora en lo digital como es la americana. Aquí está la gran batalla y no en el 5G o la guerra comercial que tan ingenuamente ha promovido alguna de las ultimas administraciones americanas, dejando al descubierto el flanco digital, auténtico campo de batalla por la hegemonía del siglo XXI.

En todo caso el balance literario de este *New Great Game* será mucho peor y no ha habido ni habrá un Kipling capaz de hacer literatura de las circunstancias de este conflicto.

III. LOS MODELOS DE DIGITALIZACIÓN

Si estamos hablando del siglo digital y del cambio social que implica, no será ocioso preguntarse cómo cada región del mundo y muy especialmente los contendientes, EEUU y China, enfocan desde el punto político y regulatorio el mundo digital.

Es una discusión trascendente en un mundo en que la digitalización afecta a todos los órdenes de la economía y la vida política y social: si la digitalización, más allá de la disrupción tecnológica es un cambio social que afecta a la economía, la sociedad y la política, el modelo de digitalización en el siglo XXI estará necesariamente muy vinculado al modelo de sociedad.

Y esto es esencial en un mundo en que la digitalización está conformando no solo la economía sino todo lo demás, desde las relaciones personales a la vida política. No podemos ser ingenuos: la manera en que se afronta la digitalización implica valores y principios y, por lo que llevamos visto, no todos los modelos de sociedad son iguales ni igualmente deseables.

Tres son los modelos a los que nos enfrentamos, el modelo americano, basado en una *light regulation*, el modelo chino, centrado en el control político del proceso digital que se configura como un cambio *top down* a modo del despotismo ilustrado, y el modelo europeo centrado en la necesidad de establecer reglas y normas en la economía digital y la digitalización.

No olvidemos que de la prevalencia de alguno de estos modelos que no son solo de digitalización sino de sociedad nos permitirá responder a la pregunta de cómo va a ser la sociedad del siglo XXI y es necesario recordar que de estos modelos solo dos están basados en los valores de la democracia representativa y la sociedad abierta.

IV. SOBREVIVIR AL TSUNAMI REGULATORIO DIGITAL EUROPEO: LAS REGLAS DEL JUEGO

Europa entró en el siglo XXI bajo el síndrome del perdedor: rezagada en el desarrollo digital y sin elementos suficientes para competir con el sector tecnológico americano, Europa se obsesiona con una pregunta equivocada ¿porque no hay un Google Europeo?

Mas adelante, en el juego de los modelos de digitalización, Europa ha creído que la única posibilidad de tener un papel en el mundo digital del siglo XXI es promoviendo un modelo europeo de digitalización combinado con el uso de la política fiscal y la de Competencia, es decir, imponiendo la "tasa Google" e imponiendo multas a Google.

En esta estrategia, incierta en su comienzo, el RGPD (Reglamento General de Protección de Datos) ha sido un inesperado éxito, al convertirse –contra todo pronóstico– en un *standard global* de facto en el tratamiento de datos a nivel mundial.

La acuñación del efecto Bruselas y la firme apuesta regulatoria de Europa como *rule setter* han basado la estrategia digital de la Unión Europea en los cinco últimos años.

Todavía es pronto, más allá del éxito del RGPD, para saber si esta estrategia va a ser un éxito, hay mensajes cruzados: por una parte, los agentes digitales están corrigiendo sus políticas más discutibles, pero por otro anuncian un día sí y otro también que determinadas innovaciones y nuevos servicios no serán implementados en Europa debido a la estricta regulación europea de los servicios y mercados digitales.

Pero lo que si podemos afirmar es que, aunque llegue a ser un éxito, una estrategia regulatoria es, en sí misma, insuficiente: el liderazgo regulatorio debe ir acompañado de un intenso desarrollo de la digitalización de la economía y la sociedad. Regresemos a la diferencia entre Economía Digital y la Digitalización explicada anteriormente: el terreno europeo es el de la Digitalización de la economía y muy especialmente de sus empresas líderes en los sectores tradicionales.

Y en este contexto no parece que la situación sea mala; si bien es cierto que Europa se encuentra rezagada en el desarrollo de la economía digital y en sectores tan críticos como la Inteligencia Artificial, el Cloud o la Computación Cuántica no lo es menos que los sectores tradicionales lideres de la economía europea (Banca, Energía, automóvil, etc.) lo son también de la digitalización.

Valga por todo el ejemplo de Inditex, una de las pocas empresas que ha sabido convertir en el comercio electrónico una estructura física en una ventaja competitiva.

Esta estrategia se ha visto inesperadamente favorecida como consecuencia del COVID 19 por numerosos fondos públicos, inimaginables hasta hace poco, que han hecho que la gran mayoría de los *Next Generation Founds* (de un total nada desdeñable de 700 mil millones de euros) se hayan dedicado a promover y acelerar la digitalización de la economía europea y sus empresas pequeñas, medianas y grandes.

Pero además, no debe olvidarse que la otra misión estratégica que Europa se ha impuesto es garantizar una digitalización basada en valores y en los valores de las sociedades democráticas occidentales que han instituido la base de nuestro modelo de progreso social.

Consecuencia de todo esto es que nuestras empresas y los operadores jurídicos se enfrentan ahora a un auténtico *tsunami* regulatorio: en los últimos 5 años la UE ha aprobado regulación que afecta de manera radical a todos y cada uno de los aspectos esenciales de la economía digital y la digitalización y lo ha hecho como hemos visto y salvo en los casos en que el Tratado Europeo no le permite, mediante Reglamentos (*Acts*) que garanticen la eficacia inmediata y homogénea de las nuevas normas.

Así en estos años se ha regulado las redes de nueva generación Gigabit, la producción de chips, la ciberseguridad (DECR,NIS 2 Y DORA), las esencialisimas DMA y DSA que regulan los servicios y mercados digitales y las dos normas de datos y gobernanza de los datos y sin olvidar la primera regulación integral en el ámbito de la Inteligencia Artificial.

Cabe destacar de estas nuevas regulaciones, además de la de la IA, toda una nueva regulaciones ex ante del mercado digital inspiradas en la regulación de las Telecomunicaciones, a través de la DSA y la DMA con grandes consecuencias a corto y medio plazo, una regulación comprensiva por primera vez con el uso de los datos en la actividad económica y un régimen de Ciberseguridad indispensable en el mundo en que vivimos.

No parece que este proceso pueda continuar en el periodo legislativo 2024/2029 con semejante intensidad y actividad las instituciones europeas deberán enfocarse más en la ejecución de lo ya regulado que en incorporar nuevas normas.

Cosa distinta es si esta profusa y a veces confusa normativa va a consolidar o no el liderazgo regulatorio europea que inauguró el RGPD y su aplicación. Este es un tema que solo el tiempo resolverá.

V. INTELIGENCIA ARTIFICIAL: LA PRÓXIMA BURBUJA

Sirva todo lo anterior como preámbulo y contexto para abordar el gran problema digital de nuestro tiempo: la Inteligencia Artificial y sus efectos no solo en la economía sino en la sociedad en su conjunto.

Podríamos empezar con una archimanida referencia al preámbulo del Manifiesto Comunista de Marx y Engels cuando dice aquello de que "un fantasma recorre Europa…".

Hoy un fantasma recorre no solo Europa sino todo el mundo y no es el comunismo, es la Inteligencia Artificial y sus peligros.

No hay día ni medio de comunicación escrito o audiovisual que no este prestando una atención inusitada a la Inteligencia Artificial, incluyendo *The Economist* que ya le ha dedicado 6 portadas, y no tanto a sus posibilidades y beneficios para el desarrollo social y económico como a sus peligros y riesgos.

También hemos visto como ha creado escándalo que un prestigioso premio fotográfico se iba a otorgar no a una fotografía auténtica sino a una creada por Inteligencia Artificial o como las noticias dan cuenta diaria de empresas sustituyen a sus trabajadores humanos por sistemas de inteligencia Artificial, huelga en Hollywood incluida.

Hoy se encuentra en el centro de una opinión pública que se plantea cuáles son los riesgos de su aplicación y si es necesario limitarla o prohibirla para evitar lo que algunos *luditas exquisitos* nos han advertido: la posibilidad de que la IA pueda destruir el Mundo o apoderarse de él, temor poco fundado que no hace sino inducir a la parte menos informada de la sociedad a albergar un miedo irracional y carente de fundamento.

¿PODRÍA LA IA SER UN RIESGO EXISTENCIAL PARA NUESTRAS SOCIEDADES Y DEMOCRACIA?

La Inteligencia Artificial está transitando aceleradamente de la utopía a la distopia y no deja de llamar la atención que en un mundo tan complejo hayan brotado como setas expertos en una materia de gran dificultad tecnológica como es la Inteligencia Artificial. En una entrevista hace varios años uno de los mayores expertos españoles en la materia, Xabier Uribe, decía en certero diagnóstico, que "en el mundo de la Inteligencia Artificial hay mucho *cantamañanas*". El paso del tiempo no ha hecho más que confirmar este diagnóstico.

Probablemente el entusiasmo e interés generalizado en la Opinión Publica por la Inteligencia Artificial se deba a la irrupción *ChatGPT*, algo que no supone en si mismo una gran novedad pero que si ha permitido por primera vez al gran público experimentar, ver y tocar la Inteligencia Artificial.

VI. EL APOCALIPSIS QUE VIENE

EL FENÓMENO DE LA INTELIGENCIA ARTIFICIAL REQUIERE UN COMENTARIO TRANQUILO Y SOSEGADO.

Vayamos por partes: todo cambio de era y la Digitalización como hemos visto lo es, desata miedos y recelos, movidos unas veces por la ignorancia y otras veces por los daños que ocasionan las innovaciones en quien vive anclado en la era anterior o no tiene posibilidad de prescindir de ella.

La Digitalización no podía ser ajena a este proceso y hemos visto como la irrupción de determinados modelos de negocio de la mano de las plataformas digitales ha suscitado la reacción de los afectados.

En ejemplo paradigmático de ello ha sido la movilización de los taxistas, nuestros lúditas más castizos, que tras mucho esfuerzo han llegado a entender que es mejor adaptarse que desaparecer y mejor influir en la elaboración de las reglas que bloquear las calles. A la postre todos hemos entendido que en la economía digitalizada hay sitios para los nuevos y los antiguos, también para los taxistas.

No han sido tampoco ajenos a este movimiento de resistencia a la digitalización los lúditas *light* por antonomasia, los sindicatos, que, más sofisticados, hacen proclamación de su fe en la tecnología y la digitalización para a continuación y por la puerta de atrás promover todo tipo de obstáculos normativos al desarrollo de la economía digital tal y como ha sucedido en la regulación del teletrabajo o de los llamados *riders*. Incluso hemos visto como en España en las últimas modificaciones legislativas en materia laboral y en el último acuerdo social los sindicatos han introducido cláusulas o disposiciones tendentes a limitar el uso de la Inteligencia Artificial.

HASTA AQUÍ NADA NUEVO, NADA QUE NO FUERA PREVISIBLE NI QUE NO HUBIÉRAMOS VISTO ANTES.

Lo que no esperábamos con la irrupción de la Inteligencia Artificial era la aparición de un grupo de *lúditas exquisitos*: expertos en Inteligencia Artificial o pensadores innovadores del siglo XXI, grupo heterogéneo muy lejos de

ser *cantamañanas*, que desde el profundo conocimiento técnico o desde la atalaya de la filosofía política nos alertan de las consecuencias negativas que para nuestra sociedad tiene y tendrá la Inteligencia Artificial, que ellos han contribuido a desarrollar o difundir y la conveniencia de plantearse seriamente su desarrollo o incluso establecer una moratoria de varios años.

Es como si Thomas Newcomen o James Watt hubieran alertado en su momento de los grandes peligros de la Industrialización y hubieran pedido la moratoria en el desarrollo de la Máquina de Vapor.

Así, hemos sabido que uno de los antiguos directivos de Google, Geoffrey Hinton, ni más ni menos uno de los creadores de las redes neuronales, ha alertado de los peligros futuros del desarrollo de la Inteligencia Artificial, anunciando que abandona Google. Para Hinton, la Inteligencia Artificial "puede llevarnos al fin de la civilización en cuestión de años" y cree que la Inteligencia Artificial va a mejorar las capacidades del cerebro humano en un plazo corto de entre 5 y 20 años. Sus palabras han tenido una repercusión extraordinaria en la opinión pública y publicada porque concretan estos miedos difusos que ya existían y los pone en boca no de un apocalíptico sino de un integrado, en la de un experto, un Pablo de Tarso en plena caída del caballo, un converso, figura que siempre tiene, como la de los pesimistas, mucho prestigio intelectual.

Por su parte Yuval Noah Harari desde su atalaya *telavivi* y la resaca de su prestigio bien ganado pero estirado hasta la saciedad también nos ha advertido, en The Economist nada menos, de que la inteligencia Artificial amenaza la supervivencia de la Civilización.Se pregunta Harari "¿qué pasará en el curso de historia cuando la Inteligencia Artificial se haga cargo de la cultura y comience a producir historias, melodías, leyes y religiones?. Herramientas anteriores como la imprenta y la radio ayudaron a difundir las ideas culturales de los humanos, pero nunca crearon nuevas ideas culturales propias. La Inteligencia Artificial es fundamentalmente diferente. La Inteligencia Artificial puede crear ideas completamente nuevas, una cultura completamente diferente". En los últimos tiempos Harari no ha hecho sino incrementar el tono de sus advertencias sobre el impacto existencial de la IA quizá en la necesidad de incrementar las ventas de su último libro, Nexus.

En resumen, un festín para los morbosos y catastrofistas de toda condición y también los defensores de la *tecnoresistencia*.

Mas equilibrados, otros pensadores desde la tecnología como Kai Fu Lee o desde el pensamiento como Michael Ignatieff se han pronunciado de manera más moderada, alertando de los problemas en materia de datos, transparencia o seguridad o llamando la atención del hecho de que los sistemas de Inteligencia

Artificial se están desarrollando a una velocidad mayor que nuestra capacidad para regularlos.

VII. ¿ESTÁN JUSTIFICADOS ESTOS TEMORES?, ¿QUÉ ES Y QUE NO ES LA INTELIGENCIA ARTIFICIAL?

Hagamos primero una revisión general sobre lo que es la Inteligencia Artificial y, como en tantas ocasiones, lo que no es porque solo considerando también una definición de tipo negativo podremos tener una idea cercana de lo fundado o no de estos temores.

La Enciclopedia Británica (que aunque nadie lo sospeche todavía existe aunque en ediciones *online*) define el concepto de Inteligencia Artificial como "la habilidad de una computadora o de un robot controlado por una computadora de realizar tareas normalmente asociadas a seres inteligentes"

Veamos otra definición, menos convencional y más detallada:

La inteligencia artificial (IA) se refiere a la capacidad de las computadoras y los sistemas para realizar tareas que, tradicionalmente, requieren inteligencia humana. La IA se enfoca en el desarrollo de algoritmos y sistemas que pueden aprender y tomar decisiones por sí mismos, basándose en datos y experiencias previas.

En general, existen dos tipos de IA: la IA débil o estrecha y la IA fuerte o general. La IA débil se enfoca en solucionar tareas específicas, como reconocimiento de voz, visión artificial o traducción de idiomas, mientras que la IA fuerte tiene como objetivo replicar la inteligencia humana en todos sus aspectos.

La IA se ha utilizado en una amplia variedad de aplicaciones, incluyendo asistentes virtuales, sistemas de recomendación, detección de fraudes y reconocimiento de patrones en imágenes médicas. La IA también ha sido objeto de controversia, ya que algunos temen que pueda reemplazar trabajos humanos o ser utilizada de manera malintencionada.

En realidad, esta descripción de la inteligencia Artificial es un autodefinido: es la descripción que nos da *Chat GPT*, es decir, la Inteligencia Artificial definida por la inteligencia Artificial.

Sigamos con *Chat GPT* y preguntémosle sobre los peligros de la IA:

La inteligencia artificial (IA) presenta ciertos riesgos y desafíos que deben ser considerados y abordados para garantizar su uso seguro y ético. Algunos de los peligros potenciales de la IA incluyen:

1. *Desplazamiento de trabajos: A medida que los sistemas de IA y automatización se vuelven más avanzados, es posible que reemplacen algunos trabajos tradicionales, lo que podría llevar a una mayor desigualdad económica.*
2. *Sesgo algorítmico: Los algoritmos de IA se basan en datos y pueden reflejar sesgos y prejuicios involuntarios. Si no se detecta y corrige, esto podría llevar a decisiones discriminatorias y injustas.*
3. *Autonomía no deseada: Si los sistemas de IA se vuelven demasiado avanzados, es posible que tomen decisiones por sí mismos sin la supervisión adecuada, lo que podría tener consecuencias no deseadas.*
4. *Dependencia tecnológica: La creciente dependencia de la IA y la tecnología podría hacer que las personas pierdan habilidades valiosas y se vuelvan menos capaces de tomar decisiones importantes sin la ayuda de la tecnología.*
5. *Uso malintencionado: Al igual que cualquier tecnología, la IA también puede ser utilizada para fines malintencionados, como la manipulación de información o la creación de armas autónomas.*

Es importante abordar estos riesgos y desafíos para garantizar que la IA sea segura y beneficiosa para la sociedad en general. Esto puede incluir la regulación gubernamental, la investigación ética y la implementación de sistemas de supervisión y transparencia.

Ninguna sorpresa, en el mar de información sobre el que navega y en el que pesca *ChatGPT* nada ha sido dicho de los peligros críticos y existenciales de la Inteligencia Artificial.

Busquemos otra definición más precisa, propuesta por la Unión Europea:

La inteligencia artificial (IA) se refiere a los sistemas que muestran un comportamiento inteligente mediante el análisis de su entorno y la toma de acciones, con cierto grado de autonomía, para lograr objetivos específicos. Los sistemas basados en la IA pueden estar puramente basados en software, actuando en el mundo virtual (por ejemplo, asistentes de voz, software de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento de voz facial y de voz) o la IA se pueden integrar en dispositivos de hardware (por ejemplo, robots avanzados, coches autónomos, drones o aplicaciones de Internet de las Cosas)".

Partiendo de esta última definición, muestra del talento de las instituciones europeas para definir las cosas, ¿cuáles son las características de la Inteligencia Artificial?

Es el mundo de los algoritmos: la Inteligencia Artificial se basa en algoritmos que extraen resultados de un gran volumen de datos y tiene una velocidad de aprendizaje y entrenamiento superior a la humana.

ADEMÁS LA INTELIGENCIA ARTIFICIAL NO SE DEGRADA Y ES INCANSABLE

Tiene un comportamiento inteligente, cierto grado de autonomía y no solo se refiere al software sino también al hardware y está directamente relacionada con el desarrollo de la Computación Cuántica.

Así, tres son los atributos de la Inteligencia Artificial: conducta Inteligente, cierto grado de autonomía y se refiere tanto al software como al hardware.

No es ocioso precisar que lo que en realidad está generando este nivel de debate y alarma no es la Inteligencia Artificial que lleva muchos años con nosotros sino la llamada Inteligencia Artificial generativa, es decir aquella que es capaz de aprender de los datos de entrada y patrones,generando contenidos con un cierto grado de novedad frente al sistema informático tradicional en el que nada sale si no entro antes, frente al cual el sistema es capaz de razonar e introducir información adicional que no se había introducido tal cual.

LA GRAN PREGUNTA: ¿EN QUÉ SE DIFERENCIA LA INTELIGENCIA ARTIFICIAL DE LA HUMANA?

Tres son las características de la Inteligencia que no puede replicar la Inteligencia Artificial: es multipropósito, el conocimiento es reutilizable y la creatividad e imaginación.

Esto nadie lo discute, la diferencia de criterio radica en es que esto es, para unos, simplemente el estado del arte a fecha de hoy y para otros la Inteligencia Artificial no es ni será nunca jamás inteligencia, es solo una sofisticada combinación de algoritmos, capacidad de procesamiento y grandes bases de datos.

Profundicemos un poco en este dilema de la mano del profesor López de Mántaras un gran experto en este tema y que ha analizado en profundidad la diferencia entre la Inteligencia Artificial y la Humana en su artículo “Algunas reflexiones sobre el presente y futuro de la Inteligencia Artificial”.

Cita a Newell y Simon cuando definen la inteligencia como un sistema de símbolos físicos como fuentes de las acciones inteligentes que tienen sustrato físico, que puede ser biológico o electrónico, neuronas o circuitos, es decir Inteligencia Artificial o Inteligencia Humana.

La Inteligencia Artificial para ser comparable debería ser de carácter general, como es la inteligencia humana: *Deep Blue* puede ganar al ajedrez a cualquier ser humano, pero no es Inteligencia Artificial de carácter general, tampoco lo es *ChatGPT*.

De ahí surge la sustancial diferencia entre Inteligencia Artificial fuerte y débil, la fuerte es la general, como la humana, aunque como dice el filósofo Searle no toda Inteligencia Artificial general (multitarea) es fuerte. El resto no tiene capacidad de desplazar ni rivalizar con ella. Solo la Inteligencia Artificial general puede tener los efectos y peligros que se atribuyen a la Inteligencia Artificial por nuestros lúditas de última generación.

Y AQUÍ NOS ENCONTRAMOS CON LA BÚSQUEDA DEL SANTO GRIAL: LA INTELIGENCIA ARTIFICIAL GENERAL FUERTE.

Hasta ahora no hemos visto más que Inteligencia Artificial débil y específica, incluyendo la cognitiva, que es al que está detrás de *ChatGPT.La* duda y el debate no está aquí, sino si en algún momento la veremos como parecen defender nuestros lúditas exquisitos.

La inmensa complejidad neuronal del cerebro hace pensar que la singularidad, la *superinteligencia* artificial, réplica del cerebro que supere a la inteligencia humana es, como dice este autor, una predicción con poco fundamento. Todavía está vigente la afirmación de Hubert Dreyfus, un filósofo, de que la Inteligencia Artificial fuerte de tipo general es tan inalcanzable como para los alquimistas del siglo XVII transformar el plomo en oro.

¿DÓNDE ESTÁ LA BARRERA INFRANQUEABLE PARA LA INTELIGENCIA ARTIFICIAL?

Esta en el cuerpo, con su capacidad de actuar con el entorno a través de las experiencias. El cuerpo y no solo el cerebro conforma la inteligencia y sin cuerpo no puede haber inteligencia de tipo general.

Tampoco las máquinas tienen sentido común ni se les puede dotar de él y sin esto no es posible una compresión profunda del lenguaje ni una interpretación

inteligente de lo que capta un sistema de percepción visual o táctil. Estos conocimientos son producto de nuestras vivencia y experiencias como humanos y dan una generalidad y profundidad inalcanzable para las máquinas: solo funcionan en entornos restringidos y preparados y no entienden la relación causa efecto, solo la correlación.

Los sistemas de Inteligencia Artificial no aprenden nada en el sentido humano: es el olvido catastrófico, en otros términos, la falta de capacidad multipropósito: lo que aprenden en un área no pueden utilizarlo deductivamente en otra.

Concluye López de Mántaras que "por muy inteligentes que lleguen a ser las futuras inteligencias artificiales, incluidas las de tipo general, nunca serán iguales a las inteligencias humanas ya que, el desarrollo mental que requiere toda inteligencia compleja depende de las interacciones con el entorno y estas interacciones dependen a su vez del cuerpo, en particular del sistema perceptivo y del sistema motor. Ello, junto al hecho de que las máquinas no seguirán procesos de socialización y culturización como los nuestros, incide todavía más en que, por muy sofisticadas que lleguen a ser, serán inteligencias distintas a las nuestras. El que sean inteligencias ajenas a la humana y, por lo tanto, ajenas a los valores y necesidades humanas nos debería hacer reflexionar sobre posibles limitaciones éticas al desarrollo de la IA"

Siendo más precisos, ¿qué es lo que nunca podrá hacer la Inteligencia Artificial?

1. *Partir de cero y crear algo que no existe.*
2. *Opinar, ser autocrítico.*
3. *Sentir emociones: La IA no puede experimentar emociones como los humanos. A*
4. *Pensar de manera abstracta y creativa o improvisar.*

No parece que con esta realidad encima de la mesa sea muy racional angustiarse por la capacidad de la Inteligencia Artificial de desplazar a la humana y poner fin a nuestra civilización. Seria contrario a la experiencia decir que esto nunca podrá suceder pero si que esto no podrá suceder en un futuro alcanzable.

VIII. LOS RIESGOS REALES DE LA INTELIGENCIA ARTIFICIAL Y SU REGULACIÓN

Parece, por tanto, que no tiene mucho sentido continuar la discusión sobre el advenimiento del Apocalipsis o las prohibiciones definitivas o moratorias temporales que es un ejercicio inútil o melancólico y centrémonos en aquello que realmente merece la pena y no dejemos que el árbol del apocalipsis tecnológico nos impida ver el bosque de los riesgos reales y abordemos los problemas concretos y no los que todavía no existen y nunca van a existir, sin regodearnos en distopias. Más nos vale enfocarnos en problemas más reales y riesgos más tangibles.

Despejado el apocalipsis ¿quiere esto decir que el uso de la Inteligencia Artificial no implica riesgos?, ¿quiere esto decir que no hay que regular la Inteligencia Artificial?

Al contrario, esta necesidad es más urgente que nunca porque como ha dicho Ignatieff no debería evolucionar la Inteligencia Artificial mucho más deprisa que nuestra capacidad para regularla.

Centrémonos pues en la necesidad de abordar la regulación de la Inteligencia Artificial y los problemas reales de su implementación que como señala Kai Fu Lee se resumen en cuatro: transparencia, datos, seguridad e imparcialidad.

También en como detectar los errores en el diseño, los dilemas éticos y cómo abordar aquellos casos en que, excepcionalmente, deba plantearse la prohibición de su uso como sucede con las armas totalmente autónomas basadas en la Inteligencia Artificial.

De lo que estamos hablando aquí, puesto en contexto, es también de modelos de digitalización, ya que la Inteligencia Artificial es una parte central de la regulación de la Economía Digital y la Digitalización.

Se necesita regulación, pero por razones más mundanas que salvar a la humanidad.

Los sistemas de IA existentes plantean preocupaciones reales sobre el sesgo, la privacidad y los derechos de propiedad intelectual. A medida que avanza la tecnología, otros problemas podrían hacerse evidentes. La clave es equilibrar la potencialidad de la IA con una evaluación de los riesgos y una regulación capaz de adaptarse a sus cambios y dinámica.

¿CUÁL ES LA SOLUCIÓN?

Nos vamos a encontrar de nuevo con los diferentes modelos de Digitalización a que nos hemos referido anteriormente, que representan diferentes modelos de sociedad y lo que es más importante valores distintos: los valores occidentales defensores de la Sociedad Abierta y los autoritarios de una China emergente.

Los modelos regulatorios responden al mismo esquema bien conocido: *light touch,* regulación intensa y control autoritario.

Nadie se sorprenderá en saber que el primero es el modelo anglosajón, sobre todo de un Reino Unido que quiere jugar fuerte en esta Industria y atraer inversiones en Inteligencia Artificial, Europa el segundo en el marco de su Estrategia Digital y China el tercero poniendo en marcha una regulación intrusiva que requiere que todos los productos desarrollados por la Inteligencia Artificial se registren y sean sometidos a un escrutinio por el gobierno antes de comercializarse y siendo sus creadores responsables del contenido, que no puede subvertir el poder del Estado, incitar a la secesión, o perturbar el orden económico y social y debe ser conforme a los valores socialistas. Más claro, agua.

Y nunca debemos olvidarnos que estos tres modelos implican modelos de Sociedad y que este es el *Great Game* del siglo XXI, ésta es la batalla por la hegemonía, la lucha por la supremacía tecnológica, la del desarrollo de la digitalización, pero sobre todo la batalla por el liderazgo en la Inteligencia Artificial y la Computación Cuantía. No es exagerado decir, aunque suene ampuloso, que quien domine la Inteligencia Artificial dominara el Mundo.

IX. LA VISIÓN EUROPEA

Merece la pena detenernos algo más en el modelo europeo que es pionero y el más desarrollado y el que, salvo en China, va a acabar influyendo a los demás reguladores de acuerdo con la vocación europea de *rule setter* digital: probablemente será la europea la regulación en cuyo espejo se miraran la mayoría de los países –y Latinoamérica– cuando se planeen la necesidad de abordar regulatoriamente la Inteligencia Artificial sin que Estados Unidos vaya a quedar totalmente fuera de esta influencia.

El Reglamento de Inteligencia Artificial (Artificial Intelligence Act) ya es una realidad legislativa y fue aprobado definitivamente el día 13 de marzo de 2024 y entró en vigor el 1 de agosto de ese mismo año, pasando después a fase de ejecución.

La regulación Europea responde una serie de regla generales jurídicas, morales y metajurídicas:

1. *Supervisión humana: debe ser supervisada por seres humanos, con las medidas de contingencia apropiadas*
2. *Robustez y seguridad: Los sistemas deben ser "resistentes" a posibles intentos de manipulación o piratería y estar equipados con planes de contingencia.*
3. *Privacidad y control de datos: La privacidad de los datos de los ciudadanos debe garantizarse durante todo el ciclo de vida de la inteligencia artificial.*
4. *Transparencia: La IA debe ser transparente, lo que significa ser capaz de reconstruir cómo y por qué se comporta de cierta manera, y aquellos que interactúan con esos sistemas deben saber que es inteligencia artificial, así como qué personas son responsables de ella.*
5. *Diversidad, no discriminación y equidad: la inteligencia artificial debe tener en cuenta la diversidad social desde su desarrollo para garantizar que los algoritmos en los que se basa no tengan sesgos discriminatorios directos o indirectos.*
6. *Bienestar social y ambiental: el desarrollo tecnológico debe tener en cuenta su impacto social y ambiental de una manera que sea sostenible y ecológicamente responsable.*
7. *Responsabilidad: La inteligencia artificial y sus resultados deben rendir cuentas a los auditores externos e internos.*

La *Artificial Intelligence Act* responde a los siguientes Principios:

- Abordar los riesgos creados específicamente por las aplicaciones de IA;
- Proponer una lista de aplicaciones de alto riesgo;
- Establecer requisitos claros para los sistemas de IA para aplicaciones de alto riesgo;
- Definir obligaciones específicas para los usuarios de IA y proveedores de aplicaciones de alto riesgo;
- Proponer una evaluación de la conformidad antes de que el sistema de IA se ponga en servicio o se ponga en el mercado;
- Proponer la aplicación después de que dicho sistema de IA se ponga en el mercado;
- Proponer una estructura de gobernanza a nivel europeo y nacional.

Bajo estos principios, la regulación europea, en proceso de ejecución, aborda la Inteligencia Artificial basándose en los riesgos específicos creados por ella, distinguiendo una pirámide de cuatro niveles de riesgo en las aplicaciones desarrolladas con su uso : riesgo inaceptable, riesgo alto, riesgo

limitado y riesgo mínimo, lo que determinara un nivel diferente de regulación: así, en las aplicaciones de alto riesgo se definen obligaciones de un análisis de conformidad con carácter previo a su lanzamiento al mercado y determinadas obligaciones que contempla el Reglamento.

No puede desconocerse tampoco el impacto que en las futuras políticas públicas digitales tendrá el Informe Draghi de septiembre de 2024, t*he future of European competitiveness*, que plantea un diagnóstico adecuado para el desarrollo futuro de la economía digital en Europa y, lo que es más importante, una nueva política de competencia más acorde a los desafíos del siglo XXI.

X. EL MODELO AMERICANO

Cuando ya habíamos asumido como verdad universal que la regulación del mundo digital era un *European affaire* debido a la vocación europea de *rule setter* y pensábamos por tanto que iba a ser el Reglamento europeo de IA la primera norma en sentar las bases regulatorias de la IA recibimos el día 30 de Octubre de 2023, via comunicado de prensa de La Casa Blanca la noticia de que el Presidente de los EEUU, Joe Biden, había decidido aprobar una *Executive Order* en materia de Inteligencia Artificial.

Antes de entrar en él, somero, análisis del contenido de la norma es necesario tener en cuenta cual es el contexto legislativo de esta disposición:

En primer término la Executive Order necesita una norma habilitante y produce efectos inmediatos sin necesidad de la aprobación del Congreso de los Estados Unidos.

Lo que hace la *Executive Order* es en buena medida impartir instrucciones a buena parte de los Departamentos/Ministerios de la Administración para que un plazo fijado por la *Executive Order,*dependiendo de los casos, de un mínimo de 90 días y un máximo de 365 se aprueben las disposiciones, planes y acciones establecidos en la propia Orden y que afecta a los Departamentos más importantes del Gobierno Americano: Defensa, Sanidad, Energía, Seguridad Nacional, Comercio y Justicia.

En cuanto a su contenido, la Orden Ejecutiva, en sus propios términos "establece nuevas normas para la seguridad de la Inteligencia Artificial, protege la privacidad de los estadounidenses, promueve la equidad y los derechos civiles, defiende a los consumidores y los trabajadores, promueve la innovación y la competencia, impulsa el liderazgo estadounidense en todo el mundo".

Esta orden forma parte de la estrategia global de la Administración para la innovación,y se basa en medidas anteriores de naturaleza bien distinta, de

manera especial la labor que condujo a compromisos voluntarios de quince empresas digitales líderes para impulsar el desarrollo seguro y fiable de la Inteligencia Artificial.

Desde este punto de vista supone un giro en la tradicional aproximación *light touch* de la regulación americana hacia terrenos de mayor intervención pública sin llegar a replicar la regulación europea, mucho más intervencionista.

Lo primero que llama la atención, frente a la regulación europea, mucho más enfocada en los riesgos de la Inteligencia Artificial y la necesidad de regular la fenomenológica de los servicios, la Orden tiene una visión más global y holística de los desafíos de la inteligencia artificial, sus riesgos pero también sus extraordinarias oportunidades.

Ocho son los ámbitos en que se desarrolla esta norma:

1. Nuevas normas para la seguridad de la IA.
2. Proteger la intimidad de los estadounidenses:
   sin salvaguardias, la Inteligencia Artificial puede poner aún más en peligro la privacidad.
3. Fomento de la equidad y los derechos civiles
4. Defender a consumidores, pacientes y estudiantes.
5. Apoyo a los trabajadores
6. Fomento de la innovación y la competencia
7. Fomento del liderazgo estadounidense en el extranjero
8. Garantizar el uso responsable de la IA por parte de los gobiernos.

Algunas de las iniciativas más relevantes que serán llevadas a la practica en los próximos meses:

Los desarrolladores de los sistemas de IA más potentes compartirán los resultados de sus pruebas de seguridad y otra información crítica con el gobierno federal estadounidense especialmente cuando supongan un riesgo grave para la seguridad nacional, la seguridad económica nacional o la salud colectiva.

Se protegerá a los estadounidenses del fraude y el engaño posibilitados por la Inteligencia Artificial estableciendo normas y mejores prácticas para detectar contenidos generados por esta y la autenticación de contenidos oficiales.

Se establecerá un programa avanzado de ciberseguridad para desarrollar herramientas de Inteligencia Artificial que permitan detectar y corregir vulnerabilidades en sistemas críticos.

Para garantizar que la IA fomente la equidad y los derechos civiles, la se deberán proporcionar orientaciones claras a los suministradores, los diseñadores programas federales de software y los contratistas federales para

evitar que los algoritmos se utilicen para exacerbar la discriminación, Abordando la discriminación algorítmica y perseguir las violaciones de los derechos civiles relacionadas con la IA.

Desarrollar principios y mejores prácticas para mitigar los daños y maximizar los beneficios de la Inteligencia Artificial para los trabajadores, abordando el desplazamiento de puestos de trabajo, las normas laborales, la equidad, la salud y la seguridad en el lugar de trabajo, y la recopilación de datos.

En el ámbito internacional, ampliar los compromisos bilaterales, multilaterales para colaborar en la Inteligencia Artificial.

Para tener un conocimiento cabal de proceso al que esta *Executive Order* da comienzo es necesario esperar un tiempo a su ejecución pero ya podemos decir dos cosas:

- Los EEUU han decidido acelerar en la regulación de la Inteligencia Artificial, haciendo frente a una demanda de la Opinión Pública y algunos sectores empresariales, incluyendo las empresas tecnológicas, que venían reclamando la necesidad de que por parte de la Administración se establecieran reglas que permitieran el desarrollo de la Inteligencia Artificial pero a la vez hicieran frente a sus riesgos.
- Frente a la regulación europea que se plasma en el Reglamento de Inteligencia Artificial que se basa en la pirámide de los servicios (prohibidos, regulados, notificados y permitidos) EEUU opta por un modelo que abraza todas las facetas de la Inteligencia Artificial, sus riesgos pero también sus oportunidades.
- Esta normativa no puede desligarse del carácter estratégico que tiene la Inteligencia Artificial y la pugna con China por su dominio.
- Es necesario resaltar que las diferencias entre la Regulación Europea y la americana ponen de manifiesto no solo las distintas estrategias regulatorias sino la diferente realidad económica de la que traen causa: EEUU es una regulación hecha a medidas de una economía digital líder que produce servicios de inteligencia Artificial; por el contrario Europa es una economía en que prevalece el uso de servicios de IA, lo que explica la preocupación regulatorias centrada en los riesgos.

XI. DESAFÍOS COMUNES

Una vez que está aprobado el reglamento de Inteligencia Artificial en Europa facilitara una tarea ineludible y necesaria geopolíticamente: la determinación de reglas comunes de Europa y Estados Unidos en la regulación y desarrollo de la Inteligencia Artificial que mantenga el siempre difícil equilibrio entre la innovación y la necesidad de afrontar los riesgos que la Inteligencia Artificial entraña para las sociedades y sobre todo para los individuos.

En este sentido la reciente cumbre sobre Seguridad de la Inteligencia Artificial en el Reino Unido, The Bletchley Declaration, así como las actividades de las Naciones Unidas UN Advisory Board tienen un papel relevante pero no debemos olvidar los límites de la cooperación internacional que exigen una colaboración estrecha entre EEUU y Europa en la promoción de un régimen regulatorio de la inteligencia Artificial que responda a los valores de las Sociedades Democráticas y el respeto a los derechos de los ciudadanos.

Descritos los rasgos principales de los modelos regulatorios habrá que recapitular sobre cuáles son los retos fundamentales de la regulación que constituyen esos desafíos comunes:

1. Es necesario encontrar el equilibrio regulatorio y evitar las tentaciones prohibicionistas o que establezcan moratorias: la regulación de la Inteligencia Artificial no debe matar su desarrollo ni frenar la innovación.
2. Hay que abordar problemas realmente existentes y no de futuro imperfecto y entre ellos el más relevante y delicado es el relativo a la compatibilidad con la GPRD y el uso de los datos.

   También es necesario garantizar la ciberseguridad y prevenir la desinformación y los *deep fakes*.
3. Una regulación pragmática de la Inteligencia Artificial debe centrarse más que en dirigir su desarrollo y establecer que es lo que se puede hacer, en lo que no se puede hacer y fijar límites que han de respetarse en el ámbito de la Inteligencia Artificial.
4. No hay algoritmos éticos o no éticos: el objeto de la regulación no pueden ser los algoritmos, el objeto debe ser las personas y las normas que deben seguirse y sobre todo evitarse en su diseño y desarrollo.
5. Es evidente el riesgo de los sesgos de todo tipo, especialmente ideológicos o de género, en el desarrollo de la Inteligencia Artificial, pero evitarlos no puede llevar a la construcción de una Inteligencia Artificial políticamente correcta que limite el desarrollo de la tecnología ni que impida que sea un cauce que también refleje la pluralidad social.

6. El CEO de Google, Sundar Pichai, ha sugerido la creación de una agencia internacional poniendo los ejemplos, muy poco afortunados, de la IATA o la IIEA. La Inteligencia Artificial es un fenómeno global y tiene que aspirar a una regulación global pero la existencia de una división radical de modelo, occidental y oriental, va a hacer esta tarea muy difícil. Durante un tiempo a lo más que podemos aspirar es que aquellos usos abiertamente peligrosos, por ejemplo, en el desarrollo de determinadas armas sea objeto de acuerdos internacionales o al menos bilaterales. De nuevo la disuasión mutua en el mundo geopolítico.
7. Es necesario regular el impacto las consecuencias del desarrollo de la Inteligencia Artificial en el empleo y en la forma de trabajar, pero desde un equilibrio que no mate la innovación.
8. En determinados casos, lo que la UE llama riesgos inaceptables, será necesario prohibir la utilización de los sistemas de Inteligencia Artificial en determinadas áreas como la de las armas autónomas.
9. No estaría de más empezar considerar riesgos asociados a la computación cuántica en relación con la seguridad de los datos o la fiabilidad a largo plazo del *blockchain*.

XII. DESAFÍOS ÉTICOS

Fuera de esta lista hay que tener en cuenta los desafíos éticos: hemos descubierto que detrás de los fríos algoritmos hay alma y como toda alma es compleja y, a veces, perturbadora.

El problema de la Inteligencia Artificial no está en el sistema, está en sus diseñadores. Toda esta regulación debe dirigirse a mantener los estándares éticos de nuestra sociedad y sus valores: libertad, igualdad y preservar la verdad en el mundo de la inteligencia Artificial.

Estos problemas de regulación vuelven a centrarnos en lo que es la esencia del problema o, mejor, de la solución: el foco fundamental de toda regulación de la Inteligencia Artificial debe ser sobre todo la transparencia: la transparencia sobre quien diseña los algoritmos, la transparencia de saber cuándo una noticia o representación digital es auténtica o *fake*, la transparencia de conocer cuáles son los límites, la transparencia de saber cómo se usan nuestros datos. Nada nuevo, desde la Revolución francesa la transparencia ha sido uno de los pilares de la Democracia y la Sociedad Abierta.

Solo así podremos acostumbrar a una sociedad digital y digitalizada de ciudadanos libres e informados a vivir de manera pacífica, responsable y

despreocupada, en el mejor sentido, y no dejarse seducir por los profetas del Apocalipsis digital.

Lo ha resumido muy bien Ignatieff: el reto es como garantizar la permanencia de la verdad.

Una vez más la resolución a los desafíos que desde el punto de vista político y regulatorio plantea la digitalización es la de la permanencia de los valores y principios de la sociedad democrática occidental en los cambios, a veces vertiginosos y radicales, provocados por los avances tecnológicos y sus consecuencias, y de cómo preservar la libertad colectiva e individual y los principios de la Sociedad Abierta.

XIII. UN EPÍLOGO LATINOAMERICANO

¿Y LATINOAMÉRICA?

Latinoamérica tiene un mismo perfil geopolítico y retos semejantes que Europa: son regiones rezagadas en la Economía Digital y fuera de la pugna geopolítica, pero con dos diferencias que reflejan la mejor posición europea:

Falta integraciónn regional y un mercado único que le permita aprovechar la escala de sus infraestructuras de Telecomunicaciones de nivel mundial.

Un menor tejido industrial competitivo a nivel global en los sectores tradicionales.

Hay que huir del pesimismo Digital Latinoamérica, arraigado hoy en la región como lo estaba a principios de siglo en Europa: siendo estos dos problemas esenciales, sin embargo Latinoamérica tiene una gran ventaja frente a Europa: población joven, numerosa y bien formada, lo cual representa también una gran oportunidad política. El desarrollo de la Digitalización depende de tanto de los conocimientos de frontera como de tener una mano de obra joven, flexible y preparada para entender el uso de los medios digitales en su actividad. Tan importantes son los ingenieros como las personas de conocimiento medio.

La otra gran fortaleza de la región es la existencia de unas infraestructuras de comunicación de nivel mundial por lo que Mantener el desarrollo de Infraestructuras de Banda Ancha como prioridad política a medio y largo plazo es esencial y debe ser la espina dorsal de la política de espectro.

Además latam ha tenido una trayectoria en la regulación de Telecomunicaciones que ha sido positiva y tiene que preservar y profundizar este legado. Y en este contexto, como en el pasado, es crucial huir del populismo regulatorio e introducir coherencia normativa.

Latinoamérica tiene la masa crítica necesaria para tener una escala competitiva en el siglo XXI pero esto no sirve de nada si no dispone de un marco económico irónico común que permita dar valor a esta escala: los mercados fragmentados por razones fiscales o regulatorias son unas losa insoportable en el desarrollo de la Digitalización de la economía en América Latina.

Aunado a ello, latinoamérica tiene que construir desde el pragmatismo un entorno regulatorio regional que le permita tener la escala suficiente para ser relevante en el mundo global del siglo XXI

Tamaño es condición necesaria pero no suficiente latinoamérica tiene que crear un Mercado Digital Único.

Esta es una gran diferencia con Europa, pero, siguiendo con los paralelismos, no debe olvidarse que la UE que hoy conocemos no empezó a andar con su parte política e institucional: Europa comienzo con unas ambiciones mucho más limitadas: el mercado común de solo dos productos: el carbón y el acero, dando lugar a la CECA sin la cual no podemos entender la actual UE.

Pues bien, hoy en el siglo XXI el Carbón y el Acero son los servicios y productos digitales y la Inteligencia Artificial.

¿CUÁLES SON LOS ASPECTOS ESENCIALES DE ESTE CAMINO AL MERCADO ÚNICO?

En primer lugar, el protagonismo activo y liderazgo de los reguladores.

¿PORQUE LOS REGULADORES?

Por qué han hecho un gran trabajo en el mundo de las Telecomunicaciones, huyendo del populismo regulatorio y facilitando que hoy las infraestructuras de telecomunicaciones sean altamente competitivas, aunque deberán acompañar a la digitalización con un mayor desarrollo.

Son o pueden ser los Reguladores del Mundo Digital. Al contrario que los reguladores europeos confinados en las barreras del sector tradicional y excluidos del diseño de la nueva regulación ex ante digital, los reguladores en Latinoamérica todavía están a tiempo de ser los reguladores del mundo digital, pero para ello tienen que abordarlo desde el punto de visto regional y no local. Ante la incapacidad de los gobiernos de encontrar marcos regulatorios comunes, esta tarea puede y debe ser abordada por los reguladores de Telecomunicaciones.

Están alejados del Debate político partidista y de los conflictos entre países.

Tiene credibilidad para promover este proceso y articular en torno suyo una coalición: empresas y Organismos de Cooperación.

Los reguladores cuentan con una estructura de trabajo en común que lleva funcionando muchos años y está bien engrasada.

En segundo lugar, contar con un mapa de ruta que permita abordar los elementos esenciales para construir un mercado único, comenzando con un análisis de las semejanzas y diferencias de las regulaciones nacionales.

Una vez hecho esto deberán abordarse, al menos, los aspectos siguientes:

- Las barreras al comercio electrónico en latam.
- Protección de los derechos de propiedad intelectual e industrial.
- Mercado único de Datos.
- Servicios Digitales y Grandes Plataformas.
- Política de Competencia regional en el ámbito digital.
- Fiscalidad única.
- Un marco único de Ciberseguridad.

XIV. LA INTELIGENCIA ARTIFICIAL

Este es un aspecto esencial y capital en el desarrollo digital y aquí el precedente europeo es muy valioso: al igual que Europa, Latinoamérica no es un productor de IA sino un usuario y sería un error dramático regular la IA partiendo de una realidad inexistente.

Desde esta perspectiva la visión semejante a la europea es necesaria: los riesgos que comporta el uso de la inteligencia artificial y la construcción de esta pirámide de riesgos, especialmente los riesgos altos son la base de este ejercicio.

Un análisis y monitorización de la aplicación de la AI Act en Europa debería o permitir a Latinoamérica realizar una importación inteligente de esta normativa, intentando descargar la regulación europea de su complejidad innecesaria, consecuencia de su complejidad institucional y la diversidad de cultura políticas y jurídicas que conviven en la UE. Descargar al RÍA de su carga burocrática será una ventaja competitiva importante para una regulación moderna de la IA.

Un último comentario: la reciente aprobación del acuerdo UE/Mercosur marca el camino o a seguir y la necesidad e afrontar los retos de una zona de libre comercio como la que se establecerá es un acicate importante para que sentar las bases del mercado Digital latinoamericano sea tan importante y urgente como realista.

Aunque el tratado no hace explícita referencia al mercado digital su impacto sobre áreas como el comercio electrónico al agua si se refiere, es muy importante y además abre un proceso de compatibilidad regulatoria entre Europa y Latinoamérica que se irá profundizando con el tiempo.

THE AGE OF AI IN CHINA: A PARTIALLY CONSIDERATION ON SOME NEGLECTED ISSUES

SUMMARY

I. Introduction **II.** A methodology: The Daoist philosophy can guide us to interpret the AI in China **III.** A paradigm and its shift: The rise and fall of a big power or many big powers **IV.** Does the AI in China verify the theory "Dong Sheng Xi Jiang"? **V.** China's response to the industrial revolution **VI.** The "great divergence" rather than the "great convergence" in AI between the US and China **VII.** China's Trying to Govern the AI **VIII.** In the parallel worlds, the different AI worlds? **IX.** Which world today's and future's China belongs to? **X.** Concluding remarks

> PANG, ZHONGYING

Dr. Pang Zhongying, a scholar in humanities and social sciences, teaches Global International Relations/International Political Economy and Global Economic History at the Sichuan University in Chengdu, China. He is currently a Visiting Senior Fellow at the Yusof Ishak Institute (Institute of Southeast Asian Studies - IEAS), Singapore. He is a former Visiting Fellow at the Brookings Institution in Washington DC, United States.

I. INTRODUCTION

In the field of artificial science and technology, I have been not literate, although I have been learning the Computer Science to catch up the change 21st century science and technology brings. However, as a senior student of Global International Relations, I was honored to accept a great invitation from Dr. Arturo Oropeza García for UNAM's global project entitled "Artificial Intelligence: A new era in the history of humankind". Arturo asked me to comment on "Artificial Intelligence in China, its Advances and Pending Challenges".

In doing this most difficult assignment, I was so encouraged by parts of Chinese publishing market, which is still keeping at least effective contacts with their global partners. China's largest publisher, the CITIC Press, introduced a series important AI books from the US market, although the current Chinese-American relations are with geopolitical tensions. The Chinese version of the Age of AI and Our Human Future written by Henry Kissinger with Eric Schmidt, Daniel Huttenlocher was introduced and published by the CITIC Press.[1] An AI visionary, Dr Fei-Fei Li, a Chinese-American from Chengdu at Stanford University in San Francisco,[2] published her best selling book *The World as I See: Curiosity, Exploration, and Discovery a the Dawn of AI* (Princeton University Press, 2023) by the CITIC Press in 2024. The CITIC Press is responsible to publish the Chinese version of Yuval Noah Harari's NEXUS: A Brief History of Information Networks from the Stone Age to AI.[3] Such books help deepen my understanding on the AI world.

1 See, [美] 亨利·基辛格[美] 埃里克·施密特[美] 丹尼尔·胡滕洛赫尔: <人工智能时代与人类未来>, translators: 胡利平 and 风君, 中信出版社, 2023.

2 https://profiles.stanford.edu/fei-fei-li

3 https://www.ynharari.com/book/nexus/

It is a great idea to have Henry Kissinger's views to help start our thinking on the subject of AI. Henry Kissinger thought deeply about the digital era. His *World Order: Reflections on the Character of Nations and the Course of History* (2014) includes a chapter entitled Technology, Equilibrium, and Human Consciousness.[4] In 2019, Kissinger wrote about the increasing tendency to give control of nuclear weapons to computers operating with AI.[5]

Kissinger often drew an analogy between nuclear weapons and AI. He travelled to Beijing on July 2023 for his final global visit in person. The People's Republic of China has been with nuclear deterrence since 1960s. Kissinger visited China more than 100 times since 1971. HK published his On China in 2010.

China's official news reports on HK's farewell journey included no any content on AI. However, according to The Economist, a British global journal, during the meeting, among the words Kissinger delivered to Chinese leader Xi Jinping, he focused in part on how to protect the world from the dangers of AI.[6]

In writing this piece, by doing literature review, I know China's top leader worries the AI's danger to the humanity[7]and Chinese diplomats call for must controlling AI.[8]

In the following, based on my specialty in International Studies in Economy, History, Philosophy and Politics, I try to give an extremely partially assessment on AI in China.

II. A METHODOLOGY: THE DAOIST PHILOSOPHY CAN GUIDE US TO INTERPRET THE AI IN CHINA

The legendary founder of Taoism, China's only indigenous philosophical cornerstone of local culture for over two thousand years, is reputed to have been Lao Tzu/Laozi (老子), who's representative work is entitled Dao De Jing.[9]

The 1st Chapter of the Dao De Jing(道德经):

4 Henry Kissinger, *World Order*, Chapter 9, 2014
5 https://en.wikipedia.org/wiki/Henry_Kissinger
6 https://www.economist.com/china/2024/08/25/is-xi-jinping-an-ai-doomer
7 https://www.economist.com/china/2024/08/25/is-xi-jinping-an-ai-doomer
8 Chinese Ambassador Fu Cong's remarks at the United Nations Security Council, https://www.mfa.gov.cn/eng/xw/zwbd/202410/t20241024_11515508.html
9 See, Wayne Ng, Finding the Way: A NOVEL OF LAO TZU， "Historical Note", Hong Kong: Earnshaw Books Ltd， 2018.

> "A Way called Way isn't the perennial/curtain Way. A name that names isn't the perennial/curtain name: the named is mother to the ten thousand things, but the unnamed is origin to all heaven and earth. In perennial/curtain Absence you see mystery, and in perennial/curtain Presence you see appearance/manifestations. Though the two are one and the same, once they arise, they differ in name. One and the same they're called dark-enigma, dark-enigma deep within dark-enigma, gateway of all mystery.(道可道 , 非恒道 ; 名可名 , 非恒名。无名 , 天地之始 ; 有名 , 万物之母。　故 , 恒无 , 欲以观其妙 ; 恒有 , 欲以观其徼。　此两者 , 同出而异名 , 同谓之玄。玄之又玄 , 众妙之门。)[10]

The topic sentence of above chapter is"In perennial Absence you see mystery, and in perennial Presence you see appearance/manifestations". I strongly think the sentence can serve as a universal guiding principle to partially know the existing world order of AI as well.In this article, I use it as my philosophical methodology to answer the assignment of "AI in China".

III. A PARADIGM AND ITS SHIFT: THE RISE AND FALL OF A BIG POWER OR MANY BIG POWERS

Many historians wrote many the rise and fall (兴衰, Xing and Shuai) of a big power or many big powers in the world. The rise and fall paradigm has been popular in Western and even Global academia. Historians and International Relations experts still like the paradigm very much. Since the late 1990s, China's rise has been a dominant topic in such Global writings including Global History and Global Political Economy.

Years ago, I purchased two Chinese versions of *Considérations sur les causes de la grandeur des Romains et de leur décadencee* (in English, *Reflections/Considerations on the Causes of the Grandeur and Decline of the Romans*) by French political philosopher, the author of The Spirit of Laws,

10 I reference two great books for better interpreting the Laozi or Lao Tze: David Hinton, *TAO TE CHING* (without the original Chinese version of the *Dao De Jing*), Berkeley: COUNTERPOINT, 2015. And, *THUS SPOKE LAOZI: A New Translation with Commentaries of DAODEJING* by Charles Q. Wu (吴千之), University of Hawaii (Honolulu) and Foreign Language Teaching and Research Press (Beijing), 2016. Dr Wu apparently was so inspired by German philosopher Friedrich Nietzsche's Thus Spoke Zarathustra (1885). It is Dr Wu's uniqueness in helping seizing the original meaning of the Laozi in Dao De Jing.

Montesquieu(1689-1748). It is <罗马盛衰原因论>.[11] An old Chinese edition of the book before China's Cultural Revolution in 1962 and a new one in 2016 before the COVID-19, the both deeply reflect the Chinese long attachment to the greatest value of a most important European intellectual work. In other words, Chinese scholars well drew lessons of the rise and fall of a big power from the ancient Rome. At least since Montesquieu, the rise and fall of a big power or an empire or the rise and fall of many big powers (comparatively) has been an endurable paradigm not only for World Historians but for many strategists in various fields.

In the great era of China's opening up and reform (late 1970s to 2010s), China introduced many important literature of humanities and social sciences which dealt with the theme of the rise and fall of a big power or many big powers. Since the late 1990s, prominently, there have been several Chinese editions of Paul Kennedy's *The Rise and Fall of the Great Powers: Economic Change and Military Conflict* from 1500 to 2000, not surprisingly, almost all Chinese editions of the book have the same Chinese title: Da Guo Xing Shuai(大国兴衰).[12]

Since the early 1990s, while the world obsessed with the rise of China, few have also explored the fall of China.[13] Currently, the retreat of the rise of China structurally and partially is seen finally. At least China's economy faces huge challenges and even crises, the opponent narrative of the rise of China: the fall of China, is publicly opposed by both the official Chinese and the supporters of the rise of China narrative both homely and internationally.[14] Officially, China still publicizes its big or strategic judgement or theory that "the East (read: China) rises and continues to rise and the West (read: the United States) declines and continues to decline" (Dong Sheng Xi Jiang , 东升西降).[15] The decline of China as

11 See, (法) 孟德斯鸠:《罗马盛衰原因论》(1951), 婉玲译(1962),许明龙译(2016), 商务印书馆.

12 I failed to check which one is the first Chinese edition of the Paul Kennedy's The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000, but I believe it was translated based on its first original English in 1990s.

13 For example, in 2001, Gordon G. Chang, *The Coming Collapse of China*, New York, Random House, 2001, 344 pp.

14 Interestingly, Singapore's Lee Kuan Yew School of Public Policy organized an event entitled "The China Shock at the End of History". Singapore's former ambassador to the the US Chan Heng Chee (former head of Political Science Department at NUS) argued the world including Singapore is now facing the 3rd "China Shock" - the falling of China's economy after the 1st and 2nd ones - totally the Rise of China at least economically since the early 1990s.

15 China's new theory of the "Dong Sheng Xi Jiang" is part of the Xi Jinping Thought. Amid the COVID-19 crisis, China's theorists in social sciences endorsed the Xi Jinping Thought by arguing that China as the today's representation of the East continues to rise and the US as the today's representation of the West continues to decline.

a narrative is clearly and even absolutely politically not correct in China. Still, no Montesquieu-like or Paul Kennedy-like frameworks to consider not only the rise but also fall of China.

In fact, take China as an example, its rise has been so complex in the past 4 decades since the late 1970s. It is a truly non-linear evolution. The fall too. The rise of China from 1979 to 1989 was stopped in 1989-1991. In 1992, just after the collapse of Soviet Union, China (read: Deng Xiaoping) relaunched its economic opening up and reform (**改革和开放**, Gai Ge and Kai Fang) towards a "Socialist Market Economy" (**社会主义市场经济**). The rise re-emerged in the whole 1990s amid globalization's popularization driven by the US-led neoliberalist economic and foreign policy. A fundamentally improved relationship between Beijing and Washington, which lasted in nearly 2 decades, Clinton-Jiang and Bush-Hu years (1992-2011), made China the "2nd largest economy" in a changed world economy: Global Economy. China was internationally accepted to be a full member of the post-Cold War international institution World Trade Organization (WTO) in 2001. In Hu Jintao years (2001-2011), China declared it has reached to a (world) great power (Da Guo, **大国**) status.[16] China joined the others in the form of the Group of the Twenty (G20) to help deal with the "Global Financial Crisis" triggered by the Wall Street during the Bush administration in 2008. The COVID-19 was tragic but unfortunately served as a turning point of history. Actually, since the end of the pandemic in late 2022, China has to deal with the big economic trouble.

IV. DOES THE AI IN CHINA VERIFY THE THEORY "DONG SHENG XI JIANG"?

The "Dong Sheng Xi Jiang" marks a China's "paradigm shift" - one power's rise and fall to the rise of a power (read: China) and fall of a power (read: the US). Simultaneously, China finishes its ideological shift from the Euro-centric or West-centric to the Sino-centric: China only sees the rise of itself and the fall of the other (read: the US).

But, the AI in China, seemingly, does not strongly verify the "Dong Sheng Xi Jiang".

The issue of AI, perhaps provides a unique window opportunity to examine the rise of China. Has China truly risen in the field of AI? Why has the AI sector

16 Zheng Bijian, "China's "Peaceful Rise" to Great-Power Status", New York and Washington: *Foreign Affairs*, 2005. 84(5):18. in addition, Bijian, Zheng. *China's Peaceful Rise: Speeches of Zheng Bijian 1997-2005*. Brookings Institution Press, 2005.

in China developed? How about private AI investment in China? If the AI sector in China is not private economy (Min Ying Jing Ji, 民营经济) driven, how about the prospects of the AI in China?

Ninghui Sun, from the Institute of Computing Technology, Chinese Academy of Sciences (CAS) and an academician at the Chinese Academy of Engineering (CAE), is regarded as a pivotal figure in AI development in China. At the Standing Committee of the National People's Congress (China's Parliament), Sun gave a lecture/hearing entitled "The Development of Artificial Intelligence and Intelligent Computing" on April 26, 2024. In the lecture, Sun's conclusion is that no doubt the rise of China's AI continues but China faces tremendous challenges and uncertainties in the field of the AI.

"It was a strategic manifesto outlining China's AI ambitions and the associated anxieties. He emphasized AI's dual-edged nature, noting its potential for groundbreaking advancements while also highlighting security risks, particularly concerning politically sensitive information.The tech rivalry with the United States was another focal point in Sun's lecture. With superior talent, foundational algorithms, and computational power, the U.S. holds a commanding lead in AI. China's current role as a fast follower is a status quo that Beijing is desperate to disrupt. U.S.-imposed restrictions on high-performance computing products and advanced semiconductor technologies are significant obstacles, argued Sun. The ban on advanced chips like the A100 and H100 stymies China's AI progress, highlighting the strategic chokehold the United States maintains over China's tech aspirations. This impact is particularly harsh for companies like Huawei and SMIC. Sun proposed that China innovate indigenously and diversify supply chains to reduce dependency on U.S. technology, echoing Beijing's call to safeguard its high-tech future.He outlined three strategic pathways for China's AI development. The first involves aligning with U.S.-led systems, a practical but restrictive path given current geopolitical tensions."[17]

I guess Sun and his team in China's AI science and technology, is not necessarily for doing the "Dong Sheng Xi Jiang" and intentionally frame the rise of AI in China according to the new paradigm. His lecture may be interpreted as several meanings (Ji Ge Yi Si, 几个意思): Sun and his team still think China's AI or

17 A text of Sun's lecture is published by China's National People's Congress as 十四届全国人大常委会专题讲座第十讲讲稿《 人工智能与智能计算的发展》,April 30, 2024, accessed at http://www.npc.gov.cn/c2/c30834/202404/t20240430_436915.html. A report on Sun's lecture, see, Felicity Hurley, "China's AI Ambitions: Prof. Sun Ninghui Highlights Strategic Blueprint", June 15, 2024, accessed at https://thefinancialanalyst.net/2024/06/15/chinas-ai-ambitions-prof-sun-ninghui-highlights-strategic-blueprint/

AI in China is rising or developing but in terms of the very old slogan since 1950s "Catching Up" (Zhui Gan，追赶)[18] to the levels of AI innovation in the US, firstly; Secondly, he fully acknowledged China's AI development benefited greatly from the originality of AI in the US; Thirdly, he indirectly predicted a scenario of AI in China including the AI competition with the US is not optimistic. These meanings are not congruent with the "Dong Sheng Xi Jiang".

V. CHINA'S RESPONSE TO THE INDUSTRIAL REVOLUTION

While, in the US, since the dawn of the 21st century, China has been defined as a leading strategic competitor and rival in many aspects.

At a White House Roundtable on U.S. Leadership in AI Infrastructure on September 12, 2024, Jensen Huang (黃仁勳), an Chinese-American from Taiwan, the founder and CEO, Nvidia,[19]re-stressed his firm argument and prediction that AI is the driving force of a new industrial revolution across the globe.[20] Huang has repeatedly addressed the "AI is the driving force of a new industrial revolution" in the globe.[21]

China's last dynasty Qing (1644-1911) in the late 19th century much belatedly had to meet the challenges from the (1st) industrial revolution, which happened in Britain in the 18th century. Qing's industrialization - Westernization Movement (洋务运动, 1860s-1900s) really made something but generally failed: the movement did not narrow "the great divergence" between then China and Europe/West.[22]

After 1949, China conducted a Soviet Union model of industrialization. The model largely did not work with huge human and natural resources costs in 1960s and 1970s. by "reform and opening up" towards a "socialist market economy

18 Sun as a generation of 1960s, still remembers a Mao Zedong slogan "Catching Up the United States" (赶上美国，Gan Shang Mei Guo) during the *Great Leap Forward*, one of the political movements before the Cultural Revolution in 1950s.

19 https://www.nvidia.com/en-us/about-nvidia/board-of-directors/jensen-huang/

20 https://www.whitehouse.gov/briefing-room/statements-releases/2024/09/12/readout-of-white-house-roundtable-on-u-s-leadership-in-ai-infrastructure/

21 Jensen Huang, A Keynote Speech at the COMPUTEX 2024, on June 2 in Taipei, Taiwan, see, https://www.nvidia.com/en-us/events/computex/

22 A global consequence of the Britain based Industrial Revolution in 18 century is the emergence of the so-called "The Great Divergence", which is suggested by the California School.Kenneth Pomeranz's The Great Divergence: China, Europe, and the Making of the Modern World Economy (2000, PUP) led to a global (including Chinese scholars) debate in the past 2 decades.

with Chinese characteristics", China seized the opportunity of the coming of the information age in 1980s and 1990s. By joining the World Trade Organization (WTO) in 2001, China was seemingly successfully towards a market forces driven economy. China becomes a major "emerging economy" in the global economy.

Especially, exaggeratedly by ignoring existing big problems in China's research and development, China is widely regarded as a scientific superpower.[23]

Is China taking advantages of the AI opportunity to succeed in the 21st Century Industrial Revolution?

With a sense of history, I have no immediate and irresponsible answer, but I have to leave this question open here: may China's record of responses and results by markets, societies and states to the previous industrial revolutions, both the failure and success - repeat in the future?

China is seen by some influential media in the US as a rising power of AI. New York based The Time100 names "Leaders", "Innovators", "thinkers", "Shapers" of AI annually. This following are from China:[24]

- Zhuang Rongwen, Director, Cyberspace Administration of China;
- Liang Rubo, CEO and co-founder, ByteDance;
- Wang Xiaochuan, Founder, Baichuan;
- Zhang Linghan, Professor at the Institute of Data Rule of Law, China University of Political Science and Law;
- Andrew Chi-Chih Yao, Tsinghua University.

Particularly, Andrew Chi-Chih Yao, 1946-, a recipient of The ACM A. M. Turing Award in 2000. Yao used the minimax theorem to prove what is now known as Yao's Principle.[25] Yao was a native Chinese born in Shanghai. He returned to China in 2004, currently a professor and the dean of Institute for Interdisciplinary Information Sciences (IIIS) at one of China's top 2 universities Tsinghua in Beijing. Yao was a naturalized U.S. citizen and worked for many years in the US

23 Many comments in the West think so. For example, **https://www.economist.com/science-and-technology/2024/06/12/china-has-become-a-scientific-superpower.** Such comments worry China will overtake the US in the future, see, Jeff Tollefson & Richard Van Noorden, "The US is the world's science superpower — but for how long?", **https://www.nature.com/articles/d41586-024-03403-4**

24 https://time.com/collection/time100-ai-2024/

25 "In recognition of his fundamental contributions to the theory of computation, including the complexity-based theory of pseudorandom number generation, cryptography, and communication complexity.", See, https://amturing.acm.org/award_winners/yao_1611524.cfm

but renounced his U.S. citizenship in 2015 and became an academician of the Chinese Academy of Sciences.[26]Yao helped create China's first school of AI at Tsinghua University in April 2024 and was appointed as its founding dean.[27]

Yao argues China's AI is already globally competitive, almost 50% talents of the global AI are trained in or from China, the country is a largest AI talents exporter in the world.[28]

Yao's strategic importance has been realized by China's political leadership. Reportedly, Yao wrote a letter to report his works to Xi Jinping. The top leader replied Yao to congratulate his 20 anniversary returnee. Xi not only praised Yao but hoped Yao continues to help lift China's home-grown and autonomy-oriented innovation of AI science and technology.[29]

This story between a laureate of the A.M Turing Award and China's top leader may imply many things strategically:

Some globally leading scholars in science and technology in China like Yao are well treated as China's "Strategic Scientists" (**战略科学家**). China develops the AI as a most strategic project like the country pursued its nuclear weapons in the late 1950s to 1960s, when two Communist states China and Soviet Union were on the brink of nuclear conflict. Beijing critically faced Soviet Union's use of nuclear weapons in the late 1960s.[30] Soviet Union obtained the nuclear weapons in 1950s after the United States in 1940s.[31] China's nuclear program got key assistance from Moscow.

In economic development, China has claimed its "system advantages" (Ti Zhi You Shi, **体制优势**). Angela Zhang argues China's authoritarian governance model can not only support the entire AI development but enable AI companies to train their models on vast amounts of personal data. In this sense, China obtains some "AI advantage" (Ren Gong Zhi Neng You Shi, **人工智能优势**).[32]

---

26 See, https://en.wikipedia.org/wiki/Andrew_Yao

27 After Tsinghua, many state-owned leading Chinese universities are establishing their schools of AI.

28 See, China's populist newspaper *Global Times under the People's Daily*, Beijing, June 29 2024, https://world.huanqiu.com/article/4IOhiaWznCs.

29 http://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/202406/t20240612_1135187.html

30 "The Sino-Soviet Border War: Why the USSR nearly nuked China", **https://www.historyanswers.co.uk/history-of-war/the-sino-soviet-border-war-why-the-ussr-nearly-nuked-china/**

31 See, https://www.britannica.com/technology/nuclear-weapon/The-Soviet-Union

32 https://www.project-syndicate.org/commentary/how-china-uae-became-artificial-intelligence-powerhouses-by-angela-huyue-zhang-2024-09

Talents and innovation capitals are the key of keys in well developing AI in China. China faces challenges from the AI investors or entrepreneurs'"Great Escape".[33]Reportedly, there is a wave of Chinese companies to the United States, particularly California's Silicon Valley to re-develop their enterprises, although they have to face the restrictions by the US government. Such Chinese have lost their confidence and opportunities in China. They believe they can take advantage of many opportunities artificial intelligence presents in the United States.[34]

VI. THE "GREAT DIVERGENCE" RATHER THAN THE "GREAT CONVERGENCE" IN AI BETWEEN THE US AND CHINA[35]

Look at A Glossary of AI terms developed by Washington DC based think tank The Brookings Institution and its research partner at Tsinghua University in Beijing. The glossary shows some not small difference between the two AI.[36]

Look at a newly release book *Nexus: A Brief History of Information Networks from the Stone Age to AI* by Yuval Noah Harari, who depicts a most dangerous future of China-US compared with yesterday's Soviet-US:

In CHAPTER 11 The Silicon Curtain: Global Empire or Global Split?, Harari's a key word is 'Silicon Curtain'. What does the "Silicon Curtain" mean? The UK-based Guardian comments that the geopolitics between China and the US is: extrapolating from the Great Firewall of China, which prevents most Chinese citizens from accessing sites such as Google and Wikipedia, Harari supposes that, in time, Chinese and US computer systems might be completely prevented from interoperating or even communicating with one another, so "ending the idea of a single shared human reality". [37]

---

33 The Great Escape, American war film, released in 1963, that was loosely based on the true story of an ambitious escape by Allied prisoners of war during World War II. See, https://www.britannica.com/topic/The-Great-Escape

34 Li **Yuan**, "Can China Tech Find a Home in Silicon Valley?",

35 The Great Divergence means the socioeconomic shift in which the Western world overcame pre-modern growth constraints and emerged during the 19th century as the most powerful and wealthy world civilizations, eclipsing previously dominant or comparable civilizations such as 19th century China. See, https://en.wikipedia.org/wiki/Great_Divergence

36 https://ciss.tsinghua.edu.cn/upload_files/atta/1725035303351_4A.pdf

37 See, "Nexus by Yuval Noah Harari review – end of days?", September 11, 2024, https://www.theguardian.com/books/article/2024/sep/11/nexus-by-yuval-noah-harari-review-the-ai-apocalypse

The term "Silicon Curtain" is not creatively coined by Harari. New generation in the 21st century may totally forget the 'Iron Curtain' (**铁幕**, Tie Mu).[38]Lindsay Gorman wrote in 2019 when the 30th anniversary of 1989, "A Silicon Curtain is descending: Technological Perils of the next 30 years".[39] I believe most of Chinese don't know the term although they have known the "new Cold War" (Xin Leng Zhan , **新冷战**). However, Silicon Curtain is already interpreted into Chinese as **硅幕**, or Gui Mu).

Can China and the US avoid dropping a 'Silicon Curtain' to pursue their technological goals without abandoning open markets and compromising global innovation?[40]

VII. CHINA'S TRYING TO GOVERN THE AI

China initiates so-called Global AI Governance (Quan Qiu Ren Gong Zhi Neng Zhi Li, **全球化人工智能治理**). What does China's Global AI Governance mean and imply?

Put Governing AI rather than developing AI first. China has no Global AI Development Initiative but issued a Global AI Governance Initiative (GAIGI) in October 2023. To China, it seems, governing AI at multi-levels, including the global levels (for example, at the United Nations), is more pressing than developing AI.

It is a new China's new initiative after its *Big Three Initiatives*. The Big Three Initiatives have already been regarded as most important parts of *Xi Jinping Thought of Foreign Policy/Diplomacy* (**习近平外交思想**): Global Development Initiative (GDI), Global Security Initiative (GSI) and Global Civilization Initiative (GCI).[41]Currently, China still not revised the Big Three Initiatives as the Bit Four Initiatives to include its GAIGI. But, GAIGI potentially becomes

38 The Soviet not only refers to the Soviet Union. It represented the System of the "International Communist Movement" controlled by Moscow. People's Republic of China was part of the system from its founding 10 years to the early 1960s. As Beijing-Moscow split lasted from late 1960s to 1980s, it seemed China was no longer part of it. But, the PRC, in its nature, after the end of the Soviet Union, is still a full Soviet and even Stalinist state.

39 https://www.gmfus.org/find-experts/lindsay-gorman

40 Marina Yue Zhan (the author of Demystifying China's Innovation Machine: Chaotic Order), "The US and China must avoid dropping a 'Silicon Curtain'", Hong Kong: **South China Morning Post**, September 10, 2024.

41 "A Global Community of Shared Future: China's Proposals and Actions" by The State Council Information Office of the People's Republic of China, September 2023, see, https://www.gov.cn/zhengce/202309/content_6906335.htm

the fourth big initiative after the big Three. The GAIGI defines "governance" as "security" together with "development" of AI[42] and calls for stakeholders at the international levels to "develop AI governance frameworks, norms and standards based on broad consensus, so as to make AI technologies more secure, reliable, controllable, and equitable."[43]

Before the issue of the GAIGI, China began to organize the World AI Conference (WAIC) in 2018.[44]The Conference is now a government's *channel to* practice its Global AI Governance Initiative but serves as an AI learning platform for many Chinese elites. The 2019 WAIC was so impressed globally because China's Jack Ma and Elon Musk conducted a maybe historic MUSK-Ma Dialogue.[45] In the intense interaction, "Elon Musk and Jack Ma disagree about AI's threat".[46]Jack Ma ignored the importance of AI and misinterpreted the AI. China's leading AI scholars and investors (most of them so young) largely praised Musk, China's official news agency Xinhua praised Ma very much. [47]

The WAIC was paused during the COVID, but the WAIC 2024 in Shanghai continued to explore Global AI Governance.

THE SHANGHAI DECLARATION ON GLOBAL AI GOVERNANCE:[48]

"With the rapid development of AI technologies, we are also facing unprecedented challenges, especially in terms of security and ethics. We underline the need to promote the development and application of AI technologies while ensuring security, reliability, controllability and fairness in the process, and encourage leveraging AI technologies to empower the development of human society. We believe that only through global cooperation and a collective effort can we realize the full potential of AI for the greater well-being of humanity."

42 A fundamental "principle" of Xi Jinping Thought on Foreign Policy/Diplomacy is "equal emphasis on development and security" (发展和安全并重).

43 https://www.fmprc.gov.cn/web/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml

44 https://www.worldaic.com.cn/wangjie?year=2019

45 https://www.youtube.com/watch?v=f3lUEnMaiAU

46 https://www.bbc.co.uk/news/technology-49508091

47 马斯克狠狠地给马云上了一课, https://user.guancha.cn/main/content?id=435375

48 https://english.news.cn/20240704/b1ca8d45f8f54ba9b954fcd969a2304f/c.html

China promotes its *Global AI Governance Initiative* at the United Nations and other inter-governmental organizations. At the General Debate of the 79th Session of the United Nations General Assembly 2024, Foreign Minister Wang Yi, who is a Member of the Political Bureau of the Chinese Communist Party Central Committee, repeated China's existing doctrine and practice of AI Governance as below:

"China is committed to taking a people-centered approach, developing AI for good, and putting equal emphasis on development and security. We are working to explore and establish widely-recognized international rules and standards. China supports the U.N.'s role as the main channel for global AI governance and is committed to strengthening international cooperation on AI capacity-building. China has put forth the *AI Capacity-Building Action Plan for Good and for All*, and is ready to make more contributions to the sound, orderly, fair and inclusive development of AI."[49]

VIII. IN THE PARALLEL WORLDS, THE DIFFERENT AI WORLDS?

In 1950s, under the worst conditionality of the Cold War between the Soviet Union and the United States, the newly created People's Republic of China in 1949 was absolutely part of the International Communist Movement world dominated by Moscow. But,the death of Joseph Stalin in 1953, the PRC-Soviet Union tensions in the late 1950s and 1960s, and the USSR-USA (and China-US) "detente" in the late 1960s and 1970s, made Mao Zedong finalize his theory of "three worlds" (三个世界) before his death in 1976. Since then China began to identify itself as "a Third World country" or "a developing country" to pursue its more "independent foreign policy" within and beyond the International Communist Movement system.

After the Cold War, China decided to re-reform and opening up to be a member of globalization driven world/international order. In 2008, in the name of One World, China hosted the One Olympics in Beijing. Currently, China's "one foot" is still on the One World. There was a Chinese delegation of the 2024 Olympic Games, Paris, France. But, China's "one foot" is already on the another

49 https://webtv.un.org/zh/asset/k1v/k1vx3nr7l1

world.[50] China has been a pillar of the BRICS since 2009, which has been seeking an alternative world order.

Recently, an American scholar in International Relations and World Affairs consider "a dividing world": "Three Worlds: the West, East and South",[51] which is so different from his previous only one liberal international or world order.

More than 10 years ago, G. John Ikenberry asserted that"the rise of China will be integrated into the existing international order". He even deeply believed China is part of the existing One World.[52]Then a Shanghai-based newspaper used an very optimistic title to cover Ikenberry's lecture trip to Beijing: "Professor John Ikenberry: China is not another world".[53]

IX. WHICH WORLD TODAY'S AND FUTURE'S CHINA BELONGS TO?

The death of Soviet Union does not mean the death of the Soviet System. During the rapid rise of China years from 1990s to 2010s, China assured and reassured the others in the international system, particularly the US, "today's China is not yesterday Soviet Union"; China is impossible to become a Soviet state. But, finally, China is emerging as a truly heir of the Soviet Union in the world.

The author of The dean of Shandong: Confessions of a minor bureaucrat at a Chinese university, "a westerner's defence of China's political system", [54]Daniel Bell,thinks "The Communist Comeback".[55] Bell's point of view is perhaps not

50 There is a Winston Churchill **statue** on the **grounds** of the British **Embassy** in Washington, **DC, the US. "**Sculpted by William McVey, the statue has one foot on British soil, the other on American, symbolizing Churchill's honorary United States citizenship, and the ties between the two nations." See, https://winstonchurchill.org/resources/in-the-media/churchill-in-the-news/historian-richard-norton-smith-on-the-churchill-statue-at-the-british-embassy-in-washington/. Here, I borrow the "one foot on a land" and "the other on a land" to describe China's current international standing.

51 G John Ikenberry, Three Worlds: the West, East and South and the competition to shape global order, *International Affairs*, Volume 100, Issue 1, January 2024, Pages 121–138.

52 G. John Ikenberry, *Liberal Leviathan: The Origins, Crisis, and Transformation of the American World Order* (Princeton: Princeton University Press, 2011, which is translated into Chinese, entitled 自由主义利维坦: 美利坚世界秩序的起源、危机和转型. On January 2014, Professor Ikenberry visited Beijing and strongly confirmed his conclusion on future China as part of the liberal order, see, **https://en.iiss.pku.edu.cn/fj/PDF/ciss_en/upload/docs/2016-09-06/doc_9281473125195.pdf.**

53 See, https://www.thepaper.cn/newsDetail_forward_1247078

54 https://www.ft.com/content/a2a8c734-49b4-4f23-b6bd-6b7f73fae250

55 Bell, Daniel. (2023) 2023. The Dean of Shandong. Princeton University Press.

accurate, but he observed a resurrection of yesterday's Soviet Union in today's China.In higher education of the ruling ideologies, China has run at least 1440 schools of Marxism until 2021 with more than 125 thousands staff professors.[56] Economically, the entire Chinese economy is no longer "market economy" but rapidly back to the centrally planed or command economy,[57] although recently, at the 3rd the Third Plenary Session of the 20th Central Committee of the CCP, the market elements with "socialist" conditionality once again was mentioned.[58]

AI, if developed over-rapidly, may accelerates the division of the dividing world. António Guterres, Secretary-General of the United Nations repeated his worry about AI at the general debate of the 79th Session of the General Assembly of the UN in New York on September 24, 2024:

"Artificial intelligence could lead to artificial divisions across the board –a Great Fracture with two internets, two markets, two economies– with every country forced to pick a side, and enormous consequences for all."[59]

António Guterres's words are not necessarily consistent with his deeds. The UNSG declined to attend the "(first global) Summit on Peace in Ukraine in Switzerland on June 15 2024 to enhance a high-level dialogue on pathways towards a comprehensive, just and lasting peace for Ukraine."[60] But, he in the capacity of the UNSG spoke in in Kazan Russia on October 24, 2024 at the 16th BRICS Summit. His "We need peace in Ukraine. A just peace in line with the UN Charter, international law and General Assembly resolutions"[61] was responded differently in the world. Ukraine angrily criticized the UNSG's meeting with Russian President Vladimir Putin, [62]who is a defendant accused of committing war crimes by the International Criminal Court (ICC).[63]

56 The figure comes from China's Ministry of Education. see, http://www.moe.gov.cn/fbh/live/2022/54301/mtbd/202203/t20220317_608432.html

57 On the term "Command economy", see, https://www.britannica.com/money/command-economy

58 https://english.news.cn/20240718/e74d931886e64878ae6e5419d19a64da/c.html

59 **António Guterres, Secretary-General of the United Nations, Remarks at the Opening of the General Debate of the Seventy-ninth Session of the General Assembly, September 24, 2024.**

60 https://www.eda.admin.ch/eda/en/fdfa/fdfa/aktuell/dossiers/konferenz-zum-frieden-ukraine/Summit-on-Peace-in-ukraine-joint-communique-on-a-peace-framework.html

61 https://www.un.org/sg/en/content/sg/statement/2024-10-24/secretary-generals-remarks-the-16th-brics-summit-delivered

62 https://news.sky.com/story/antonio-guterres-why-is-the-un-secretary-general-in-russia-incurring-ukraines-wrath-13240299

63 https://www.icc-cpi.int/defendant/vladimir-vladimirovich-putin

The case of China in development and governance of AI may strengthen the UN SG's concern and anxiety over the AI as a threat to human ontology in the exiting international order. In facts and trends, China has transformed itself as a representative of a new world: a politically unreformed Soviet-nature system with Chinese characteristics, a foreign policy to seek a multi-polar world order together with today's Russia, an intranet hard to connect with the global-nets, a dominant state-owned economy, a controlled society with the brainwashing as education for the next generations.

What are historic and strategic implications of China's new direction towards a new world for China's AI science and technology? Look at the future, do we witness the continuity of China's rise or do we see the end of the rise of China in the field of AI?

X. CONCLUDING REMARKS

By examining some neglected but important events, cases, figures and stories, this chapter looks at partially the progresses, problems and prospects of the AI in China.

I introduce one of China's current official theory entitled "the East rises and the West declines" to examine the AI in China, which has seemingly not demonstrated the theory.

Paul Samuelson (1919-2009), the author of a classic textbook Economics: an Introductory Analysis in 1948, an American economist who was awarded the Nobel Prize in Economic Science in 1970 for his fundamental contributions to nearly all branches of economic theory,[64] who predicted the Soviet Union economy would overtake the US in 1980s.[65] But, his prediction did not happen. The "Dong Sheng Xi Jiang" is not a Samuelson-like scientific research conclusion but a political ideology. The rise of the AI in China has been due to the rise of the AI in the US. This is an international result of mutually promoted interdependence. The primacy of the AI in the US is largely acknowledged by China's leading AI scientists, who have had connections with their American colleagues. The rise of AI in China not means the the decline of the AI in the US. AI development in China faces its sustainability challenge as China deviates truly market economy and re-embraces the Soviet Union centrally planed or command economy.

64 See, https://www.britannica.com/money/Paul-Samuelson

65 From The Vintage News, https://www.thevintagenews.com/2018/08/23/soviet-gnp/

From China-US relations perspective, AI is so globally geostrategic. Currently, in the field of AI, the great divergence rather than the great convergence between China and the US is happening. In the "Age of AI", the globe is not an integrated and united world. In the future, the globe includes at least two worlds: China and the US are in different worlds.

THE IMPACT OF AI ON CHINA'S ECONOMIC DEVELOPMENT. ADVANCES AND CHALLENGES TO BE ADRESSED

SUMMARY

I. Introduction **II.** The Current State of China's Economy **III.** Main Advantages of AI Application in China **IV.** Challenges of AI Application in China **V.** Policy Recommendations **VI.** References

> JIN SONG
PhD in Economics, Senior Fellow, Director of Development Research Center, Institute of World Economics and Politics (IWEP), Chinese Academy of Social Sciences

I. INTRODUCTION

The unprecedented advancements in science and technology result profound transformations in the global political and economic landscape. A new wave of technological revolution is sweeping the world as evidence that the emerging fields such as artificial intelligence (AI), blockchain, and biotechnology are becoming pivotal engines of economic growth. AI, a branch of computer science, aims to simulate and extend human intelligence through technologies such as machine learning, natural language processing, and computer vision, and stands out as a transformative force with the potential to redefine industries and reshape economies. In recent years, significant breakthroughs in deep learning, natural language generation models (e.g., GPT), and autonomous systems have led the widespread application of AI in various fields such as healthcare, finance, energy, transportation, etc., which proves that AI is increasingly integrated into economic and societal operations and offes innovative solutions and driving efficiency.

In no exception, China's economic development in the global transformations is also at a historic crossroads. Having achieved decades of high-speed growth, China is now transitioning to high-quality development, which emphasizes innovation, sustainability, and inclusivity as the pillars of the progress. The Chinese government's focus on fostering "new quality productivity" aligns closely with the imperatives of the transition. It highlights the integration of advanced technologies like AI into economic systems to accelerate industrial upgrading, optimize supply chains, and build a modern, innovation-driven industrial framework. Besides, demographic changes in China further underscore the urgency of the transition. As the population growth slows and the labor force ages, the traditional model of growth driven by abundant low-cost labor faces

diminishing returns. It necessitates a shift toward productivity-led growth fueled by technological innovation. AI, as a key enabler of such productivity, could plays a leading role in addressing the challenges of an aging population, optimizing resource allocation, and enhancing economic competitiveness. Despite the fact that China's innovation-driven development strategy has already positioned the country as one of the global leaders in artificial intelligence research and application, the rapid integration of AI into economic systems also brings critical challenges such as ethical and regulatory concerns, and the impartiality to AI benefits. Responding effectively to these challenges will be core issues to China's high-quality development in the AI era.

This paper analyzes the primary advantages of artificial intelligence in China's economic transformation, lists the core challenges that need to be addressed and attempt to provide a systematic perspective on how AI can contribute to the high-quality development of China's economy.

II. THE CURRENT STATE OF CHINA'S ECONOMY

China's economy has entered a new phase of high-quality development after decades of rapid growth, and the economic operation now presents a series of new characteristics and challenges.

FIRST, ECONOMIC GROWTH RATE HAS DECLINED

The pace of China's economic growth has gradually slowed down from double-digit rates to a moderate-high range. Since the 2020s, the economic growth rate has noticeably decelerated with the gradual disappearance of the demographic dividend, weakening external demand, and the deepening of domestic economic structural adjustments. Though a slowdown is a normal phenomenon as an economy matures, the decelerating trend reveals deeper contradictions and issues such as insufficient domestic demand, slowing productivity growth, and uneven technological innovation capabilities. Additionally, uncertainties and complexities in the international macroeconomic environment, including geopolitical conflicts and global supply chain realignments, add further pressure on the growth rate. It necessitates that China should focus more on the development quality rather than the speed of economic growth to ensure a long-term sustainable development through optimizing resource allocation, strengthening innovation-driven strategies, and improving total factor productivity.

SECOND, TRADITIONAL INDUSTRIES NEED TRANSFORMATION AND UPGRADING

China's traditional industrial system, centered around manufacturing, has been driven by low-cost labor, resource inputs, and scale expansion. But with the adjustment of the international division of labor and changes in the stage of domestic economic development, the limitations of this model have become increasingly apparent. First, traditional manufacturing faces overcapacity. In certain industries, low-end and homogenized competition lead to resource wastage and declining profit margins. Second, the rise of emerging economies in mid-to-low-end manufacturing has intensified international competitive pressures. Meanwhile, the global economy is accelerating towards digitalization and greener practices, forcing traditional industries to leap towards higher-end, smarter, and greener operations through technological upgrades and structural adjustments. This not only involves nurturing high-tech industries represented by artificial intelligence, big data, and cloud computing, but also includes promoting the deep integration of manufacturing and services, establishing a modern industrial system powered by technological innovation.

THIRD, CHINA'S POPULATION AGING ISSUE HAS INTENSIFIED, AND THE PROPORTION OF THE WORKING-AGE POPULATION CONTINUES TO DECLINE

Statistics show that since 2012, the total number of China's working-age population has been decreasing, and it is projected that by 2035, the elderly will account for over 30% of the total population. Population aging impacts the economy in multiple ways. The decrease in the labor force directly weakens the potential momentum for economic growth, the pressure of pension payments and increased medical expenses further exacerbate the public fiscal burden, and aging significantly influences consumption patterns, with rising demands for healthcare and eldercare services, while growth in traditional consumer sectors slows. To address the challenges of aging, it is essential to increase labor productivity, to delay the retirement age, to expand the supply of skilled technical personnel, and to optimize the social security system to mitigate its impact on the economy.

FOURTH, THE LABOR SKILL STRUCTURE HAS BEEN SIGNIFICANTLY TRANSFORMED

In the past few years, the overall university enrollment rate in China has shown an upward trend, gradually stabilizing at a high level. Since the implementation of the higher education expansion policy in 1999, the gross enrollment ratio (GER) in tertiary education has experienced a significant increase, rising from 30% in 2012 to 59.6% in 2022. The underlying drivers behind the soaring rate in higher education enrollment include robust support from educational policies and a strong inclination of households to invest in education. China has consistently promoted the development of education, striving to enhance the accessibility of education and narrow the gaps between urban and rural areas, as well as among different regions. Simultaneously, with the ongoing economic transformation and technological advancements, there has been an increasing demand for highly skilled labor in society. Higher education is widely recognized as a crucial pathway to enhancing individual competitiveness. Given the high value placed on education by Chinese families, an increasing number of households are willing to provide support for their children to pursue higher education. All these have led to a notable shift in the skill structure of the labor supply.

FIFTH, CHINA FACES SIGNIFICANT PRESSURE FOR CARBON REDUCTION

Being one of the main carbon emitters in the world, China confronts significant pressures to reduce its carbon footprint. To achieve the goals of carbon peak and carbon neutrality, it is required to advance the quality of its economic development. Nonetheless, the endeavor to reduce carbon emissions is beset by numerous challenges. Firstly, the predominance of fossil fuels in China's energy mix, coupled with the relatively low share of clean energy, presents a formidable barrier. Secondly, the transition of traditional, high-energy-consuming, and high-emission industries is impeded by both technological hurdles and employment. The fact that these sectors often serve as major employers, where rapid reform could precipitate substantial job displacement. Furthermore, disparities in economic development and resource endowments across regions complicate the uniform implementation of carbon reduction objectives. To foster a green development trajectory, making the best use of technology progress is imperative for achieving a balance between economic growth and environmental sustainability.

III. MAIN ADVANTAGES OF AI APPLICATION IN CHINA

AI could play a significant role in China's economic development. Since the global economy has been characterized by sluggish performance, many countries urgently explore new sources or drives of growth. China has introduced the concept of "new quality productive forces", aiming to leverage the latest technological achievements, including AI, to promote economic development. Owing to its powerful capabilities in cross-domain data processing and analysis, AI technology can break down traditional barriers between fields, promote interdisciplinary integration, and spur innovative activities. It facilitates the research and development of new products, the improvement of modern services, and the innovation of business models, and gives rise to new economic opportunities. These innovative outcomes by AI not only significantly impact existing industries through creative destruction, transforming and elevating them, but also continuously stimulate economic expansion, fostering a new leap in social productivity and the overall high-quality development of the economy.

1. CONTRIBUTION TO INDUSTRIAL UPGRADING AND THE CONSTRUCTION OF A MODERN INDUSTRIAL SYSTEM

A modern industrial system requires the coordinated development of multiple sectors, including manufacturing and services, to achieve diverse and sustainable drivers of economic growth. In the past several years, Chinese traditional manufacturing and service industries have faced some major constraints due to a combination of internal inefficiencies and external pressures. The advent of AI has change the circumstances and it played a pivotal role in integrating and innovating these industries, driving the optimization and upgrading of industrial structures. AI has strengthened both internal advancements within industries and synergistic interactions across sectors, thereby enhancing overall economic efficiency, resilience, and competitiveness.

In the manufacturing sector, AI has significantly advanced automation and intelligence, fostering innovation, improving product quality, and enhancing delivery efficiency. These developments have enabled manufacturers to meet customized demands better from consumers, thereby increasing competitiveness in their products. AI empowers technological breakthroughs and provides positive influences, from production to resource allocation, in almost every

stage of the manufacturing process. Current research indicates that up to 45% paid work can now be automated through existing technologies, which facilitate real-time monitoring and adjustment of production lines. This leads to substantial improvements in product and service quality while reducing costs (Chui et al. 2016; World Economic Forum 2023). Moreover, AI enhances product delivery systems by analyzing large datasets to streamline distribution processes, significantly lowering operational costs, accelerating capital accumulation, and directing investments toward innovative areas. The shift encourages higher investment in research and development, thereby lays the foundation for sustained economic growth and expansion (Arntz et al. 2016; Brynjolfsson and McAfee 2014).

In the service sector, the integration of digitalization and intelligent technologies has enabled more efficient, accessible, and sophisticated service with improved quality and efficiency. For example, the labor substitution rate in the retail has increased from 53% to 60%, from 43% to 66% in finance and insurance services. Similarly, production and service models have shifted in healthcare and education. Across different industries, the service sector is providing more diversified support in high-quality to manufacturing and bringing new opportunities for Chinese manufacture. Correspondingly, manufacturing with great advancements offers new pathways for technological breakthroughs in services, creating a feedback loop that drives further mutual growth. This bidirectional synergy between manufacturing and services optimizes the structure of the labor market, generates higher-quality employment opportunities, and strongly supports sustainable economic development.

2. THE ENHANCEMENT OF LABOR PRODUCTIVITY

AI is reshaping the dynamics of labor productivity by complementing human efforts in the workplace. The synergy between AI and human labor has the potential to significantly boost efficiency, to enhance service quality, and to lower operational costs. In the context of China's economic transformation, where labor-intensive industries are giving way to knowledge- and technology-intensive sectors, AI represents a critical tool for achieving sustainable productivity gains.

First, AI helps to enhance efficiency through data processing and analysis.

AI improves operational efficiency across industries because it is able to process vast amounts of data and deliver real-time analytical insights significantly.

For example, in healthcare sector, AI helps in diagnostic processes and drug development. The AI-powered medical imaging systems have halved the time doctors spend on image analysis, reducing misdiagnosis rates and increasing the overall efficiency of healthcare delivery. In the financial sector, domestic banks have achieved up to 90% automation in transactions through AI simply because it allows financial professionals to focus only on complex investment decisions while minimizing operational costs. These advancements illustrate how does AI support routine tasks and how does it optimize high-value decision-making processes.

Second, AI leads to cost reduction through process automation.

AI excels in automating repetitive and time-consuming tasks, significantly lowering labor costs in various sectors. For instance, in customer service, AI-powered chatbots handle a vast number of routine inquiries, providing round-the-clock support while reserving human intervention for more complex issues. Major Chinese e-commerce platforms have adopted AI chatbots, cutting customer service costs by 30% while simultaneously improving customer satisfaction rates. This combination of cost savings and enhanced service quality highlights the transformative potential of AI in service-oriented industries.

Third, AI mitigates labor risks and increasing operational resilience.

The integration of AI and robotics further reduces labor risks and enhances productivity in high-risk and high-intensity environments. AI-powered robots are increasingly deployed in industries characterized by hazardous conditions, such as mining and chemical production, as well as in sectors requiring precision and endurance. These robots can operate continuously without fatigue, reducing workers' exposure to dangerous tasks and significantly lowering operational costs. Furthermore, the adoption of robotics minimizes labor disputes and reduces expenditures on welfare benefits and workplace safety, making it an attractive option for industries under cost and regulatory pressures.

Fourth, AI addresses China's demographic challenges.

AI's potential to replace and augment labor becomes increasingly valuable. Automation of routine and repetitive tasks helps mitigate the economic challenges posed by a shrinking labor pool. AI also complements human labor by enabling older workers to remain productive through assistive technologies, thus partially offsetting the demographic pressures on the labor market (Acemoglu and Restrepo 2018).

3. CATALYST FOR INNOVATION

AI stands out for its ability to simplify complex tasks and process vast datasets, a capability that distinguishes it from traditional approaches to innovation. Unlike human cognition, AI can search and analyze large-scale datasets within a remarkably short timeframe, leveraging its unparalleled computational power and expansive knowledge base. Such way of working enables AI to detect complex patterns through advanced techniques such as deep learning and predictive analytics, facilitating accurate forecasting, real-time decision-making, and an overall acceleration of the innovation process. Additionally, AI's robust cross-domain data processing capabilities break down the traditional disciplinary boundaries and foster interdisciplinary integration. The integration drives the development of new products, enhances modern services, and supports innovative business models, generating significant economic opportunities (Brynjolfsson and McAfee 2014). These outcomes from AI disrupt existing industries through mechanisms of creative destruction and stimulate continuous economic expansion. For instance, productivity gains driven by AI often lower the costs of goods and services, increasing real purchasing power and reshaping consumer demand elasticities, thereby contributing to economic growth (Bessen 2018).

AI significantly improves the model and efficiency of education and skills development of the talents, a cornerstone of innovation. Particularly in China, where rapid technological advancements are reshaping industries, traditional education systems often struggle with rigid curricula and inefficiencies, which is comparatively not conductive to labors' skills in rapidly evolving industrial demands. AI has significantly improved education and skill development through adaptive learning technologies. By personalizing learning content and methods, AI addresses individual learning needs, enhances skill acquisition and facilitates dynamic adaptation to new technological requirements. For example, AI-driven learning platforms in China are helping workers develop critical skills aligned with the demands of an innovation-driven economy. This capability elevates labor productivity and creates a robust talent pipeline to support sustained innovative activity (OECD 2019).

4. STIMULATING CONSUMPTION AND ENHANCES CONSUMER WELFARE

AI largely improves production efficiency and resource allocation and leads to enhanced product and service quality and lower prices, which are critical in

stimulating domestic demand. China's recent declining consumption results from a combination of structural and short-term factors. From the income perspective, slower growth in household incomes and widening income disparities have weakened consumption capacity and marginal propensity to consume. From the demographic perspective, accelerating aging and declining birth rates have altered consumption structures: a rising proportion of low-consumption elderly groups dampening overall consumption. Additionally, uneven urban-rural and regional development has constrained the release of potential consumption demand, while insufficient quality and alignment in product and service supply have limited the optimization and upgrading of consumption structures.

In general, AI reduces operating costs for businesses, drives down prices for products and services under competitive market conditions. It also optimizes supply chains and delivery experiences, increases consumer surplus and enhances overall welfare. The dynamic mechanism may even stimulate higher levels of consumption (Bessen 2018). The healthcare sector serves as a typical example. AI-driven technological advancements are profoundly transforming diagnostic and treatment methods, improving patient outcomes while effectively lowering medical costs and easing financial burdens, thereby enhancing welfare (Topol 2019). AI-enabled predictive analytics helps hospitals optimize resource allocation, reduce hospital stays, and improve patient welfare through multiple channels. In addressing the challenges posed by declining consumption demand in China's aging society, AI-powered health and eldercare services can meet the high-quality service needs of older populations, expand market participation among marginal consumers, and drive overall consumption growth.

Moreover, AI leverages sophisticated algorithms to analyze and interpret massive amounts of user data and then offer precise insights into consumer preferences and behaviors. This capability enables businesses to provide highly personalized products and services, significantly enhancing customer satisfaction and further improving consumer welfare (Davenport 2018; Deng et al. 2016). Technological innovation driven by AI presents opportunities for enhancing local product design and service quality, revitalizing domestic manufacturing and service industries, and partially substituting imported products and services, thus fostering domestic economic development. Through intelligent logistics, e-commerce platforms, and optimized allocation of public service resources, AI wonderfully boosts economic vitality and consumption capacity in underdeveloped regions..

5. ENHANCING RESOURCE EFFICIENCY, FACILITATING EMISSIONS REDUCTION, AND SUPPORTING ENVIRONMENTAL SUSTAINABILITY

AI also serves in a critical position in improving resource efficiency, reducing emissions, and promoting sustainable development. Its capabilities in predictive analytics, real-time monitoring, and resource optimization significantly enhance efficiency in key sectors such as transportation, manufacturing, construction, energy, and agriculture. In transportation, intelligent systems optimize traffic flow, and reduce fuel consumption and greenhouse gas emissions. In manufacturing, AI improves resource utilization by optimizing supply chains and production processes. Smart grids and AI-driven building energy management systems reduce energy waste. In agriculture, AI's precision minimizes water and fertilizer usage, and promote the sustainable practices in agricultural production. (World Economic Forum 2018; Chui et al. 2016).

For China, AI is particularly significant in advancing sustainable development. The country has committed to achieving carbon neutrality by 2060 and addressing environmental challenges, with a phased goal of peaking carbon emissions by 2030. These national objectives align closely with AI's potential to support the green transformation of traditional industries. AI-driven innovations, such as smart grids and green technologies like electric vehicles, contribute to reducing resource consumption, lowering pollution, and improving energy efficiency. These advancements support China's transition to a high-quality growth model that balances economic development with ecological conservation.

AI enables China to harmonize economic growth with environmental protection and avails the country advantageously in the global competition for sustainable development by promoting green technologies and circular economy practices. In the meantime, these efforts enhance societal well-being, demonstrating the dual benefits of AI-driven innovation in fostering economic prosperity and ecological sustainability.

IV. CHALLENGES OF AI APPLICATION IN CHINA

The widespread application of AI deeply reshapes the labor market in downsides, such as employment inequality, skill mismatches, income disparities, and governance issues. These challenges are adventurous for businesses, individuals, and policymakers. Unlike influences from any previous technological revolutions, the impacts of AI are faster, broader, and more complex. China without exception

also faces competitive risks stemming from global power rivalries though it leverages AI-driven advancements to foster economic development.

1. PROFOUND TRANSFORMATIONS IN THE LABOR MARKET

The adoption of AI is redefining skill requirements. In accordance with the latest report by the World Economic Forum, from 2023 to 2027, globally approximately 23% of jobs would be affected by technological disruption, involving 69 million new jobs and 83 million eliminated positions, which will result the profound changes in occupational structures across industries (World Economic Forum 2023).

AI-related positions, such as machine learning experts, robotics engineers, and business intelligence analysts, are in high demand that advanced skills like data analysis, machine learning, and algorithm development are urgently needed. Conversely, the existing clerical roles, such as bank tellers, postal workers, and ticket agents, are rapidly disappearing. Workers in traditional roles often face skill difficulties in meeting new demands, and the consequential mismatches between skill supply and demand appear,, which further require structural adjustments driven by automation and AI.

The paradox is that businesses cannot recruit their expected highly skilled talents, which consequently hampers the pace of AI deployment and innovation, and a large number of workers, who are not equipped with necessary skills, are forced out of the labor market and suffering from distress to reemployment. The mismatch is exacerbated by the slow adaptation of education and training systems, intensifies the labor market imbalances and leaves many workers in impotence in the evolving job demands.

AI's impact concentrates in technology-intensive industries and geographically works in relevant regions though its effects are broadly felt across sectors and locations due to its algorithm –and software-based nature. The productivity gains from AI also vary widely among firms, and the imbalance production returns exacerbate the inequalities of employment distribution. Capital –and technology-intensive companies achieve remarkable productivity improvements through AI-driven supply chain optimization and process innovation, but labor-intensive and low-tech firms often lack the resources and capabilities to integrate AI effectively, so they sometimes got stranded to stagnant or declining productivity and reduced competitiveness.

The productivity gap, often leading to widespread closures, enables high-tech firms to capture larger market shares, while low-tech firms face survival

challenges. High-tech enterprises experience surging demand for skilled talent, whereas traditional industries reliant on low-skilled labor face severe unemployment issues. Such polarization highlights the uneven distribution of opportunities and risks within the labor market.

Additionally, the distribution of displaced and created jobs is uneven across labor groups, potentially leading to significant structural unemployment. AI primarily targets cognitive tasks, including both routine tasks, such as customer support and data entry, and non-routine tasks, such as financial forecasting, medical diagnosis, and legal consulting (Webb 2020). By altering task structures, AI substitutes some human tasks while creating new tasks requiring different skills (Trajtenberg 2018).

Following employment disruptions, many workers are likely to enter informal employment. Some may face temporary unemployment and require retraining to secure new jobs, while others with weaker adaptability may experience long-term unemployment or even exit the labor market altogether. These outcomes depend on workers' attributes and the degree of technological disruption within their industries (Chui et al. 2016).

2. POTENTIAL RISKS OF INCOME INEQUALITY

Even though AI and intelligent robotics significantly enhance labor productivity, their impact on labor income growth is more complex. Unlike traditional automation, which predominantly affects low-income, low-skill workers, AI's influence on income is multifaceted. It disrupts jobs ranging from simple, routine tasks like data entry (typically middle-income white-collar roles) to high-income professions such as doctors, designers, and financial analysts (Trajtenberg 2018). This suggests that AI may disproportionately impact middle- and high-income workers.

Studies indicate that wage inequality may decrease during this process (Webb 2020). However, as some middle- and high-income workers shift to lower-skill jobs during reemployment, the labor oversupply also depresses low-skill workers' wages, suppressing overall income growth. The final distribution of labor income depends on the reemployment outcomes of displaced workers. It is a key policy challenge to ensure the workers share equally in the dividends of AI technology to eventually achieve equitable income distribution.

Regarding specific job-level income changes, AI has created numerous new technology-related roles and could create new high-income groups, such as AI specialists, who may temporarily lead to wage growth in some sectors. Current

high-income groups, such as doctors and financial analysts, must re-adapt their job roles or employment models in response to AI advancements to avoid temporary income declines. Displaced workers unfortunately may face direct income losses.

During the period of significant employment transition, retained workers may find themselves in disadvantageous position during labor-capital negotiations and have to take the downward wage pressures. In the long run, labor supply will gradually adjust in line with the new skills and job demands, and would form a convergence in wage advantages for newly created roles.

Overall, the introduction of conventional robotics shifts income from labor to capital, while AI and intelligent robotics may enable wealth accumulation for those with access to capital, data, innovative algorithms, and specialized expertise. Nevertheless, automation could lower the costs of goods and services, potentially enhance the effective purchasing power of existing incomes if cost reductions are passed on to consumers effectively.

3. COMPETITIVE RISKS ARISING FROM GREAT POWER RIVALRIES

Over the past several years, technological competition among major powers was fierce and insensitive and becomes more intricate and strategic. Higher levels of innovation and technological sophistication are critical for industrial advancement and enhance national competitiveness despite increasingly insurmountable technological barriers in developing countries. To secure a competitive edge in global industries, developed nations such as the United States, European Union members, and Japan have implemented a range of technology and industrial development strategies. These strategies position technological innovation and digitalization as the driving forces for future industrial development, and emphasizes the cultivation and attraction of highly skilled and qualified talent and fostering advancements in AI, the Internet of Things, and cloud computing. Developed nations leverage their mature innovation ecosystems and higher education institutions to strategically enhance the cultivation and attraction of top talent, particularly in science, technology, engineering, and mathematics (STEM). In 2020, over half of the PhD students in the US STEM programs were international students, with 73% of them staying in the United States after graduation. High retention rates among students who stay in the US are from China and India. For example, nearly 90% of Indian students earning STEM doctorates in 2020 stayed in the US. Following the 2018 US-China trade

frictions, the United States further intensified economic decoupling measures, including restricting technology exchanges, safeguarding critical technologies and sensitive data, and enhancing technology controls through patents, intellectual property protections, and market entry barriers. Developed nations also impose stringent market access standards and certification requirements to limit developing countries' products and services in their markets. Export controls and restrictions on technology transfer further inhibit the flow of critical technologies to developing countries. These measures exacerbate global technological inequalities, hinder industrial upgrading in developing nations, and reduce their participation in global value chains.

AI advancements compound these challenges, significantly shrinking the labor cost advantages of developing countries. AI, in conjunction with automation, reduces the demand for human labor, lowers production costs, and shifts capital investments toward innovation and efficiency improvements in manufacturing and services. These dynamics diminish developed nations' reliance on low-skilled labor from developing countries, and result in reduced investment and slower industrial upgrading in those nations. Furthermore, developed countries' monopolistic control over high-tech industries limits the ability of developing countries to transition to higher-value-added industries, and stalls employment shifts toward higher-skilled roles. Without domestic innovation capabilities, developing countries may over-rely on imported AI technologies, which will cost additional expenses and increasing production vulnerabilities, potentially lead themselves to technology-related debts. As the primary contributors to global population growth, developing countries face unique challenges under an international landscape marked by development asymmetries and monopolized opportunities. Population surpluses may fail to contribute to economic potential. Instead, it may cause widespread unemployment, deprivation of economic and educational opportunities, overconsumption of resources, and regional instability. Domestically, high unemployment, social security pressures, and governance challenges escalate fiscal burdens And internationally the diminished economic and technological leverage of developing countries further erodes their voice in global affairs (Manyika and Sneader 2018; Reijnders and Vries 2018).

4. SOCIAL AND ETHICAL CHALLENGES

AI holds immense potential to drive economic and social development, yet several challenges must be addressed before its positive impacts can be fully

realized. From a technological perspective, AI systems still require significant refinement. Errors in analysis results during reinforcement and deep learning processes hinder reliability and stability, which are critical for industrial applications. The discrepancy slows the pace of AI adoption relative to technological advancements. A key controversy surrounding deep learning lies in the opacity of its outcomes. Application scenarios such as healthcare demand transparency in the mechanisms underlying AI-generated outputs. The lack of clarity imposes a constraint on the pace of AI application as it is crucial to ensure its development does not outstrip human understanding.

AI also raises significant social and ethical concerns, which can limit its adoption. Beyond structural impacts on employment, AI's discriminatory effects may arise from its operational mechanisms. AI systems heavily rely on available data for reinforcement learning, and biases in such data can amplify algorithmic prejudice, exceeding that of human biases. This distorts production, service delivery, and resource allocation. Unchecked AI applications, such as generating fake content, pose threats to economic and social stability. Ethical principles guiding AI operations, as well as its interaction with humans, remain contentious. On a technical level, safeguarding data privacy is a major challenge, with data breaches posing risks to public safety and providing opportunities for criminal activities. Establishing robust regulatory frameworks to manage AI development and application is essential to address these issues (Brynjolfsson and McAfee 2014; Acemoglu and Restrepo 2019; Topol 2019).

V. POLICY RECOMMENDATIONS

AI serves as a critical driver for China's economic advancement, reshaping the modern industrial ecosystem. To maximize AI's positive impacts while mitigating its challenges, China must establish a comprehensive system that accelerates industrial upgrading and ensures efficient talent cultivation and allocation.

1. ESTABLISHING AN AI-COMPATIBLE TALENT STRUCTURE

Firstly, cultivating diverse talent profiles is important. On one hand, China needs leading talents with innovative thinking, entrepreneurial spirit, and global vision, as well as interdisciplinary talent capable of integrating advanced technologies and achieving "new combinations" of production factors for profit growth. Such individuals play pivotal roles in adapting to dynamic markets, driving industry innovation, and expanding global market presence. On the other hand, it is

essential to have ordinary workers with solid technical skills and digital literacy form the foundation for AI integration. Workers must maintain competitiveness, master their roles while developing skills in AI and data analysis to enhance efficiency, adapt to technological changes, and transition to tech-intensive jobs.

Secondly, education system reforms are needed. Traditional education in China appears some defects in flexibility and efficiency in responding to rapidly evolving labor market needs. There are two directions that the education system could be improved. One is innovative talent development: Shifting educational objectives from knowledge transmission to fostering creativity, critical thinking, and problem-solving skills; creating an open, interactive education ecosystem emphasizing interdisciplinary learning to break traditional academic silos; Incorporating cutting-edge technology trends and societal needs into curricula, and using AI tools for personalized learning experiences and progress monitoring. The other is collaborative education mechanisms: promoting dynamic collaboration among government, enterprises, schools, training institutions, scientific representatives, and families; aligning education with future industry trends and emerging occupations, and encouraging lifelong learning through policy incentives like subsidies to ensure th workforce adaptability to evolving industry demands.

2. ESTABLISHING A MARKET ENVIRONMENT FACILITATIVE OF AI DEVELOPMENT AND APPLICATION

It is important to cultivate a competitive market environment too. A competitive market environment for AI includes four components.

Firstly, establishing a healthy market ecosystem that can benefit countries to adapt to the AI era.

China boasts a large and mature market for the development and application of AI technologies, supported by a robust digital infrastructure – a unique advantage not available in many other countries. Efforts should be made to further cultivate this market towards maturity and healthiness. It involves opening up and jointly building an environment conducive to international cooperation in AI technology, which includes platforms that promote mutual benefits and win-win outcomes among nations. It can create a relaxed international exchange environment by facilitating international exchanges of experiences, technologies, and talents through diversified channels such as talent exchange programs, cross-border research platforms, remote work systems, and international academic conferences. Regular interaction between domestic and international talents,

which is thought to be contributive to a prosper global innovation ecosystem, will form a global network of talent exchange characterized by mutual benefit.

Secondly, ensuring orderly market competition during the process of transition.

Enterprises mastering AI and robotics technologies gain first-mover advantages, which can reshape the competitive landscape within industries. Given the high entry barriers associated with these technologies, small and medium-sized enterprises (SMEs) may find it challenging to enter the market, potentially leading to increased industry concentration. To maintain a competitive market structure, it is imperative to strengthen antitrust regulation, particularly focusing on enterprises with core AI and robotics technologies, especially leading firms and major internet companies, to prevent abuse of dominant market positions. Support for SMEs' development is also crucial. Providing policy support to assist SMEs in acquiring funds, technologies, and talents can be adopted to enhance their competitiveness. Promoting open sharing of technologies can encourage technological innovation and application. Nevertheless, the government should ensure smooth mechanisms for enterprise exit, avoiding excessive local government intervention or protection that leads to resource misallocation, particularly unnecessary subsidies for inefficient manufacturing sectors, to prevent hindrance to industrial upgrading and labor reallocation.

Thirdly, enhancing labor market flexibility and resilience and mitigating the challenges brought by AI to employment.

Addressing changes in employment structures brought about by new forms of productivity requires enhancing the flexibility and resilience of the labor market. Technological innovations and industrial upgrades will likely increase volatility in the job market and raise the proportion of informal employment and posing short-term or long-term occupational transition challenges for some workers. Employment management strategies need to balance maintaining employment stability within ensuring healthy market development and avoiding the accumulation of ineffective employment. Concurrently, efforts should be made to build proactive employment support and social security systems. Measures such as optimizing the market environment by accelerating the construction of a unified, open, and orderly national market, reducing administrative barriers, lowering enterprise operating costs, strengthening property rights and intellectual property protection, and enhancing the attractiveness of the business environment can indirectly promote the increase and improvement of employment opportunities. And means of improving the functionality of employment information service platforms by broadening channels for employment information release

and offering customized re-employment training services can accelerate skill upgrades and job matching for workers, mitigating adverse effects of frictional unemployment.

Fourthly, the refinement of a dynamic and adaptive social security system is essential.

China has already established a comprehensive and well-organized social protection network. However, in the era of rapid advancements in AI, it is imperative to bolster the system's dynamism and precision to effectively support diverse population segments. It helps individuals navigate through the challenges posed by AI, ensures that the system remains responsive to the evolving socio-economic landscape, and thereby support the healthy development of new forms of productivity. The system should smooth transition mechanisms and facilitate the seamless transition of affected workers into new roles or industries. The process includes providing personalized career counseling, retraining programs, and vocational education opportunities tailored to individual needs, ensuring that displaced workers can quickly adapt to emerging job markets. Proactive measures should be taken to reduce the impact on vulnerable groups, and then to prevent social unrest. Also, by protecting workers' livelihoods during transitions, it encourages greater willingness to embrace new technologies and participate in lifelong learning, which is critical for sustaining long-term economic growth.

VI. REFERENCES

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* 33(2): 3—30.

Acemoglu, Daron, and Pascual Restrepo. 2018. "Artificial Intelligence, Automation, and Work." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Arntz, M., Terry Gregory, and Ulrich Zierahn. 2016. "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis." OECD Social, Employment and Migration Working Papers, No. 189. Paris: OECD Publishing.

Bessen, James. 2018. "Artificial Intelligence and Jobs: The Role of Demand." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.

Chui, Michael, James Manyika, and Mehdi Miremadi. 2016. "Where Machines Could Replace Humans—and Where They Can't (Yet)." *McKinsey Quarterly*.

Davenport, Thomas H. 2018. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. The MIT Press.

Deng, Yue, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. 2016. "Deep Direct Reinforcement Learning for Financial Signal Representation and Trading." *IEEE transactions on neural networks and learning systems 28(3): 653-64*.

Manyika, James, and Kevin Sneader. 2018. "AI, Automation, and the Future of Work: Ten Things to Solve For." McKinsey Global Institute. https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for.

OECD, 2019. *Education at a Glance 2019: OECD Indicators*. OECD Publishing.

Reijnders, Laurens S., and Gaaitzen J. de Vries. 2018. "Technology, Offshoring and the Rise of Non-Routine Jobs." *Journal of Development Economics* 135: 412–432.

Topol, Eric J. 2019. "High-Performance Medicine: The Convergence of Human and Artificial Intelligence." *Nature Medicine* 25 (1): 44–56.

Trajtenberg, Manuel. 2018. "AI as the Next GPT: A Political-Economy Perspective." National Bureau of Economic Research Working Paper No. 24245. National Bureau of Economic Research. https://doi.org/10.3386/w24245.

Webb, S., 2020, "The Impact of Artificial Intelligence on the Labor Market." https://web.stanford.edu/~mww/webb jmp.pdf.

World Economic Forum, 2018. *Future of Jobs Report 2018*. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2018/.

World Economic Forum, 2023. *Future of Jobs Report 2023*. World Economic Forum. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf

ADVANCED GENERATIVE ARTIFICIAL INTELLIGENCE AND AI GOVERNANCE IN JAPAN

SUMMARY

I. Introduction **II.** The Rise of LLM Generation AI and Multimodal AI **III.** Hiroshima AI Process **IV.** Biden Harris Presidential Decree **V.** AI User Guidelines by Japan Government **VI.** Consideration of more advanced AI governance and AI legislation **VII.** Toward a New Vision of Social Development **VIII.** Bibliography

> OSAMU SUDOH
Ph.D. and Professor, Chuo University

I. INTRODUCTION

This paper reviews the technological trends in generative and multimodal AI, which are bringing about rapid changes in social systems and will be a paradigm shift that will mark an epoch in human history. Based on this, the significance of the "Hiroshima AI Process" proposed and approved at the G7 Summit 2023, which Japan chaired, as a trend that attempts to govern cutting-edge AI, and the Biden-Harris Presidential Decree (October 30, 2023) of the United States that supports this process will be discussed. The significance and challenges of the "AI User Guidelines," which were adopted as policy in Japan to implement these guidelines in a more concrete manner, and the basic policy of the AI legislation currently under consideration will be described. In doing so, we will describe the progress of Japan's international contributions to advanced generative and multimodal AI.

II. THE RISE OF LLM GENERATION AI AND MULTIMODAL AI

GPT3.5, a large-scale model for natural language processing generated by Open AI, was released in November 2022, and GPT4 was incorporated into Microsoft's search engine Bing this year; Bing is now able to write natural language prompts for Dall-e generated images.

Meanwhile, Google released Bard, a generative natural language large-scale data model, and PaLM2 was released on May 10, 2023.

Since then, the trends surrounding generative AI have had a tremendous impact on people to this day. AI as a superior tool (optimization tool) that optimizes for a given objective using a given set of means has changed dramatically. Major changes have been prepared since around 2018 at leading American companies and research and educational institutions such as Google,

OpenAI, Microsoft, Meta, IBM, Stanford University, and MIT (Massachusetts Institute of Technology) (1).

One of these was the Generative Pre-Transformer (GPT), the core of which was the development of the Transformer (2).

Transformer is a five-stage natural language processing mechanism that processes natural language, understands context, and generates words through large data sets and high-speed computer processing. The first stage consists of "tokenization of words," the second stage "embedding by vector transformation," the third stage "positional encoding," the fourth stage "transformer blocking," and the fifth stage "softmaxing using probability. These processes understand the context and generate the words that should be displayed next. In addition, a convolutional neural network developed by Google is used for the generative image processing (see Fig. 1).

FIG.1 Five elements of a transformer

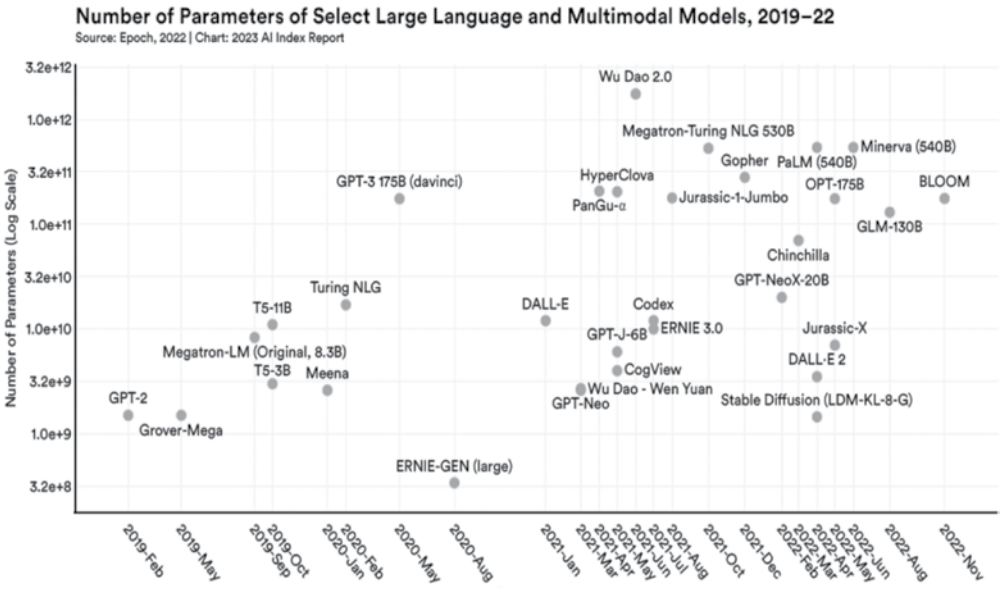
- Tokenization
- Embedding
- Positional encoding
- Transformer block
- Softmax

The text is generated through the above five steps of operation.
Mathematical processing with vector databases, context dependence and probability distributions are the main features.

This process requires a large database and fast computing resources. GPT-3, released by Open AI in the spring of 2020, has 175 billion parameters. The improved GPT-3.5, which appeared in November 2022, is also estimated to have 175 billion parameters. Even more astonishing is Google's transformer PaLM, which has 5400 parameters, and Minerva, which performs mathematical inference, which has 5400 parameters (see Fig. 2). Incidentally, Google unveiled its fine-tuned PalM-2 at Google I/O 2023 in May 2023. Gemini1.0, an early model of integrated multimodal AI R&D for mathematical reasoning and analysis of chemical formulas, DNA sequences, etc., was also released in 2023 (3).

In Japan, the Ministry of Economy, Trade and Industry has a well-known supercomputer called ABCI, which is currently being augmented. Noteworthy trends in the U.S. are those in the enhancement of mathematical and logical reasoning capabilities. For example, Minerva, a generative AI for research that was at a high level of capability as of spring 2023, is an AI that specializes in computational logical reasoning in mathematics, physics, pharmacology, and chemistry, but this was an artificial intelligence capable of running 540 billion parameters.

FIG.2 Parameters of LLM Generative Models



[SOURCE] Stanford University Human Centered AI Institute ed. Artificial Intelligence Index Report 2023, p.60

The parameters of GPT-4, which have not been made public, are said to be in excess of 1 trillion according to American researchers, and some estimate that GPT-5, which is said to be in the study phase by 2024, will be in excess of 4 trillion. The computing power of Japan's supercomputers is currently unsurpassed.

We believe that the construction and strategic use of Retrieval Augmented Generation (RAG) will become increasingly important in the future, The technology extracts information from a variety of knowledge sources that are not publicly available on the Internet and generates text from them. In the next step, the "Generator Model" uses this information to generate answers to questions and text.

After 2022, LLM-generated AI using huge supercomputers is on the rise, but recently Retrieval Augmented Generation (RAG) has been attracting a lot of attention. Instead of collecting a large amount of data from all over the world to make judgments or write sentences, for example, a RAG can process a university's confidential information, such as management, training and education, know-how, and personal information of each faculty member, using only that information. With such a volume of information, there is no need to use a high-performance supercomputer like GPT4, and RAG will be able to produce

documents with high efficiency and quality. It will also be able to evaluate reports. If this happens, we can have artificial intelligence do all the evaluation of reports on behalf of the teacher. In Japan, not only IT companies, but also many manufacturing companies, financial institutions, and local governments are focusing on RAG construction and operation (4).

With RAG, Japanese companies do not need a supercomputer and can operate with a computer of a certain level of capability, and can process important information without exposing it to the outside world. Cloud computing allows for thorough management of personal and confidential information, processing can be done in a closed environment, and copyright management is easy.

Microsoft, AWS, Oracle, Mistral, and others are also focusing on RAG. In Japan, NTT Data, Hitachi, Fujitsu, NEC, and many other IT companies are focusing on RAG development and operation. However, technically, it is not an easy task, as it requires the competence to create advanced vector databases. The way to create vector database attributes and other information is important (see Fig. 4).

FIG.3 What is Retrieval Augmented Generation (RAG)?

RAG is a widely used and important technology in natural language processing (NLP). This technology extracts relevant information from large knowledge sources and generates text from it.

The RAG consists of two components: the <retrieval model> and the <generator model>.

>The "Retrieval Model" searches for relevant information from a large amount of vector data for a given query (an imperative in DB). The Generator Model then uses this information to generate an answer or sentence to the question.

✓ GPT4 services using closed organizational data are expanding: MS, Google, AWS, Oracle, Meta, Anthropic, Mistral, NTT Data, Hitachi, Fujitsu, NEC, etc.

With the utilization of this RAG, advanced AI services can be provided for the strategic use of generative AI inside local governments and for strategic management inside corporate organizations. In this sense, RAGs are likely to steadily have a significant impact on governments, municipalities, and other administrative agencies, financial institutions, and other corporate management LLM generative AI and multimodal AI, especially RAGs, will have a significant impact on human creative and strategic actions.

Then there is the RAG system, which uses the RAG database in a cozy way and can run in conjunction with a huge LLM-generated AI system. However, if you do not manage copyrights well, you will face copyright issues when information that is supposed to be in a closed system goes out to an outside LLM, so you need to be very careful.

Furthermore, NTT and Sakana AI are working together to create a new AI. It is called MoE (Mixture of Experts). Each company has training data in its area of expertise, and the MoE, a coalition of SLM-generated AI, is an attempt to build and operate a highly capable generative AI through the cooperation of companies and research institutions that do not have the financial and human resources to gather such resources.

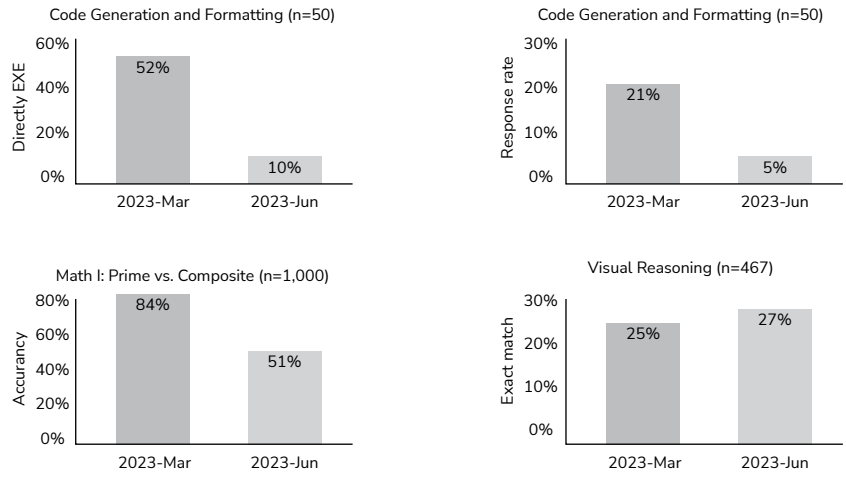
A. GENERATION AI CAPABILITIES

Wall Street Journal editors and Wall Street Journal columnists in the U.S. have compared the document production capabilities of high-profile generative AI (5).

For health information, ChatGPT is superior, followed by Google's Gemini, then Perplexity. For financial information, Gemini is the best, followed by Google's Gemini, then Perplexity. For food menus and cooking, ChatGPT is the best, and Claude, an LLM-generating AI released by Anthropic, is the best at writing work-related documents. Microsoft's Copilot is the best at creative writing, Perplexity is the best at summarizing, and Perplexity is the best at current events. In terms of coding, Perplexity was rated the highest, and ChatGPT was rated the fastest in terms of response time. Finally, from the user's perspective, the overall evaluation was Perplexity, ChatGPT, Gemini, Claude, and Copilot, in that order.

The next evaluation indicator is an interesting one (see Fig. 4).

FIG.4 Output change of GPT4.
Changes in LLM Performance Over Time



SOURCE: Lingjiao Chen, Matei Zaharia, James Zou, How is ChatGPT's Behavior Changing Over Time?

The research presented here is the work of researchers at Stanford University and UC Berkeley. It has been shown that the performance of certain commonly available LLMs can vary significantly over time.

For example, the June 2023 version of the GPT-4 was 42 percentage points lower than the March 2023 version in code generation, from 52% to 10% (n=50); in answering subtle questions, from 21% to 5%, 16 percentage points lower (n=100); in mathematical questions, from 84% to 51%, 33 points worse (n=1000). In visual reasoning, they improved by 2 points, from 25% to 27% (n=467).

The reason for this change is not clear, but may be a complex combination of various factors, including prompts, vector databases, feature settings, and token processing (6).

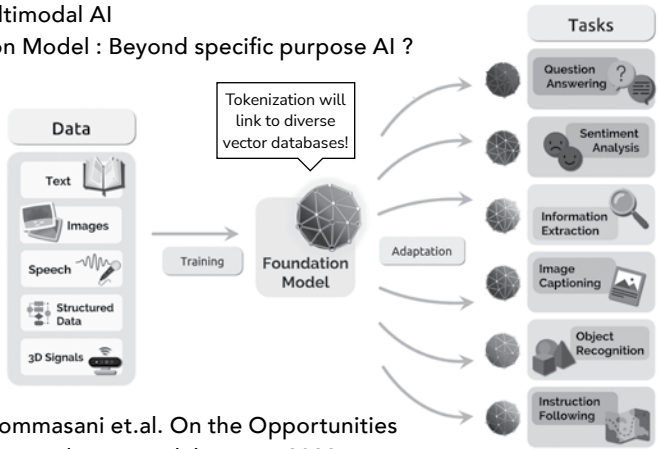
B. EVOLUTION OF GENERATION CAUSING MAJOR SOCIAL CHANGE

What has happened so far is called the democratization of AI or the popularization of AI. Until now, if you wanted to use AI, you had to learn advanced skills, the very thing you needed to be an expert in. However, with the advent of generative AI, anyone can write prompts in ordinary everyday language and use artificial intelligence to extract advanced knowledge. Hence, it is called the democratization of AI.

In the future, advanced generative AI is likely to bring enormous changes to society. And in the near future AGI (general-purpose AI) could be born.

There is a study that is highly significant in the perspective of AGI. Researchers at Stanford University in the U.S. have written a paper that describes multimodal generative AI as a foundation model. This AI is an artificial intelligence that can handle various forms of data, including language data (semi-structured data), voice data, images, videos, and structured data (numeric data) (see Fig. 5) (7).

FIG.5 Multimodal AI
Foundation Model : Beyond specific purpose AI ?



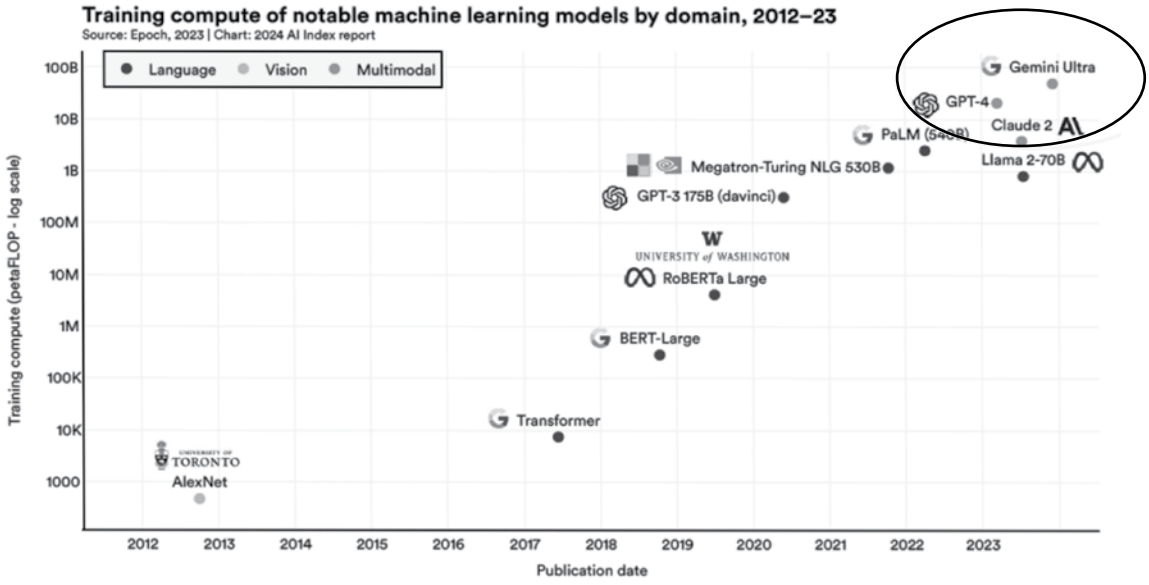
A foundation model can centralize the information from all the data from various modalities. downstream tasks.

SOURCE: Rishi Bommasani et.al. On the Opportunities and Risks of Foundation Models, arXiv, 2022

In October 2023, then US President Biden emphasized FM, or foundation model, as the most important AI.

Running a multimodal AI requires a high-performance supercomputer; see Fig. 6. If you look at this chart, the horizontal is a time series and the vertical is a scale of computation speed per second, with the highest value being petaflops x 100 billion, and the graph and scale are in logarithmic notation. The light blue color is the computational power required for multimodal AI. Compared to this, the power of ordinary LLM-generating AI, which is a hot topic in Japan, is at a much lower level.

FIG.6 Required Computational Power for Multimodal AI.

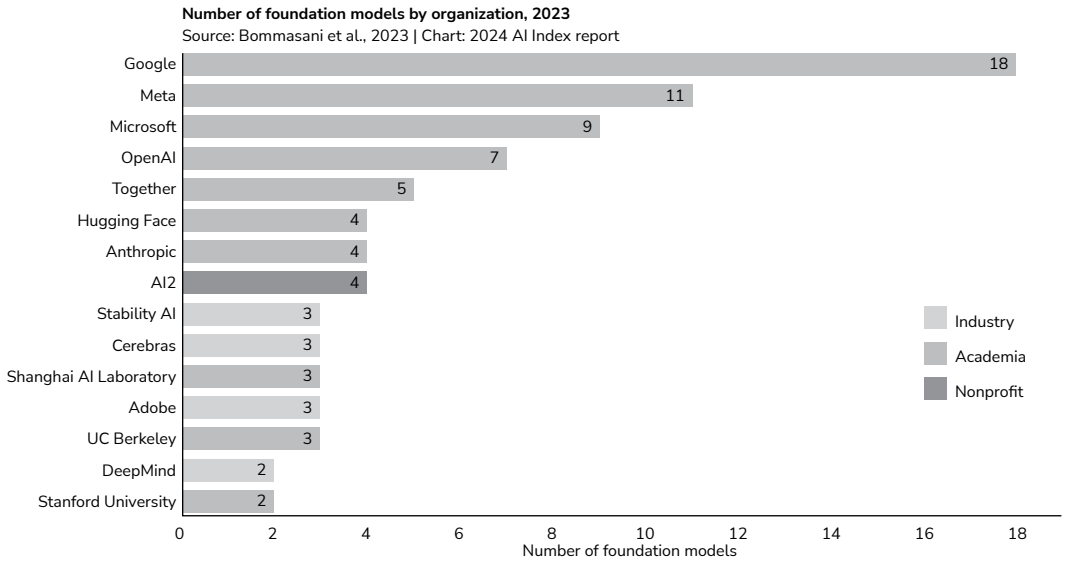


SOURCE: Stanford University Human Centered AI Institute ed. Artificial Intelligence Index Report 2024, p.25

The computational power required for multimodal AI, on which Google, Open AI Inc. and Anthropic are focusing their efforts, is in the order of awesome muscle; as of February 2025, Japanese companies and research institutions will be at a level that cannot be followed.

See Exhibit 7a. Institutions developing multimodal AI include Google, Meta, Microsoft, OpenAI, Together, Hugging Face, Anthropic, Shanghai AI Laboratory, UC Berkeley, and Stanford University.

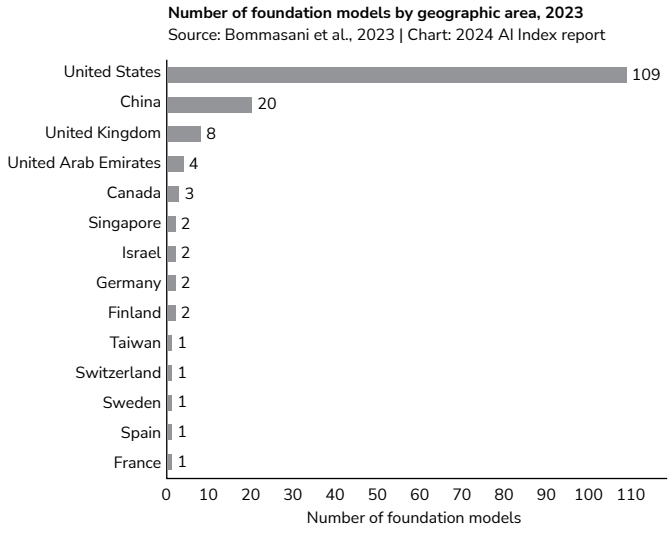
FIG.7A Development status of multimodal AI.



SOURCE: Stanford University Human Centered AI Institute ed. Artificial Intelligence Index Report 2024, p.33

And by region, where are the strongest; see Fig. 7b. The United States is by far the strongest. China is next, followed by the United Kingdom and the United Arab Emirates. Canada, Singapore, Israel, Germany, Finland, Taiwan, Switzerland, Sweden, Spain, and France follow, but Japan's name does not appear. In other words, we must say that is quite a backward country in multimodal AI R&D (8).

FIG.7B Comparison of Multimodal AI by Region.



SOURCE:
Stanford University Human Centered AI Institute ed. Artificial Intelligence Index Report 2024, P35

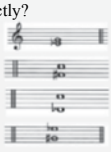
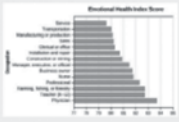
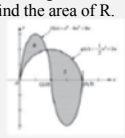
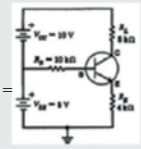
C. POTENTIAL OF MULTIMODAL AI

Google's demonstration at GoogleI/O2024 featured a radical change in the field of education using Gemini 2.0, a multimodal AI system that is still in the prototype stage, but seeks to flexibly teach the "universal law of gravity" according to the academic level of students and pupils. Students and pupils of various academic levels can be accommodated, with high level classes for high level students, and college level studies for high school and junior high school students. For students who are not good at studying, they can include a game element and use games to increase concentration and interest in teaching. This would make them more responsive than human teachers, so it is easy to imagine that quite drastic changes could occur in the field of education. There is a growing possibility that a paradigm shift could occur not only in education, but also in healthcare, business organizations, government agencies, and the social system as a whole.

In the United States, various benchmarks for measuring the logical reasoning ability of advanced generative AI have been developed and are being used in practice. MMMU is presented here as an example (see Fig. 8).

FIG.8 AI evaluation benchmark MMMU
Measuring the General Reasoning Ability of Multimodal AI MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI

Sample MMMU questions
Source: Yue et al., 2023

Art & Design	Business	Science
Question: Among the following harmonic intervals, which one is constructed incorrectly? **Options:** (A) Major third <*image 1*> (B) Diminished fifth <*image 2*> **(C) Minor seventh <*image 3*>** (D) Diminished sixth <*image 4*> 	**Question:** ... The graph shown is compiled from data collected by Gallup <*image 1*>. Find the probability that the selected Emotional Health Index Score is between 80.5 and 82_ **Options:** (A) 0 (B) 0.2142 **(C) 0.3571** (D) 0.5 	**Question:** <*image 1*> The region bounded by the graph as shown above. Choose an integral expression that can be used to find the area of R. **Options:** **(A)** $\int_{0}^{1.5} [f(x) - g(x)]dx$ (B) $\int_{1.5}^{0} [g(x) - f(x)]dx$ (C) $\int_{0}^{2} [f(x) - g(x)]dx$ (D) $\int_{0}^{\theta} [g(x) - x(x)]dx$ 
Subject: Music; **Subfield:** Music; **Image Type:** Sheet Music; **Difficulty:** Medium	**Subject:** Marketing; **Subfield:** Market Research; **Image Type:** Plots and Charts; **Difficulty:** Medium	**Subject:** Math; **Subfield:** Calculus; **Image Type:** Mathematical Notations; **Difficulty:** Easy
Health & Medicine	**Humanities & Social Science**	**Tech & Engineering**
Question: Youe are shown subtraction <*image 1*>, T2 weighted <*image 2*> and T1 weighted axial <*image 3*> from a screening breast MRI. Whats is the etiology of the finding in the left breast? **Options:** (A) Susceptibility artifact (B) Hematoma **(C) Fact necrosis** (D) Silicone granuloma 	**Question:** In the political cartoon, the United States is seen as fulfilling which of the following roles? <*image 1*> **Options:** (A) Opressor (B) Imperialist **(C) Savior** (D) Isolationist 	**Question:** Find the VCE for the circuit shown in <*image 1*>, Neglect VBE **Answer: 3.75** **Explanation:** ...IE = [(VEE) / (RE)] = [(5 V) / (4 k-ohm)] = 1.25 mA; VCE = VCC - IERL = 10 V - (1.25 mA) 5 k-ohm; VCE = 10V - 6.25 V = 3.75 V 
Subject: Clinical Medicine; **Subfield:** Clinical Radiology; **Image Type:** Body Scans: MRI, CT.; **Difficulty: Hard**	**Subject:** History; **Subfield:** Modern History; **Image Type:** Comics and Cartoons; **Difficulty:** Easy	**Subject:** Electronics; **Subfield:** Analog electronics; **Image Type:** Diagrams; **Difficulty: Hard**

SOURCE: Stanford University Human Centered AI Institute ed. Artificial Intelligence Index Report 2024, p.113

MMMU is an acronym for A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI MMMU is a measure of reasoning ability at the college level of acquisition, It is designed to measure overall competence in Art & Design, Business, Science, Health & Medicine, Humanity & Social Science, and Tech & Engineering. Various other benchmarks have been developed, but so far each has its own strengths and weaknesses and is still in the trial-and-error stage(9).

Entering the year 2024, the multimodal AI Gemini 2.0 was reportedly compared to human logical reasoning ability. According to the report, Gemini achieved a score of 90.0%, outperforming human experts in 30 of the 32 academic benchmarks used in LLM R&D (note that Open AI's O2 also reportedly surpassed human capabilities in the winter of 2024).

Remarkably, Alpha Fold3, a multimodal AI advancement by Google's DeepMind published in the scientific journal Nature in May 2024, had a huge impact on the state of medicine. It is now able to simulate protein 3D structures in a short period of time and predict hierarchical interactions and binding between DNA, mRNA, and ligands. It is highly anticipated that the detailed elucidation of the mechanisms of protein interactions and binding will revolutionize the research and development of drugs for the treatment of cancer and other diseases.

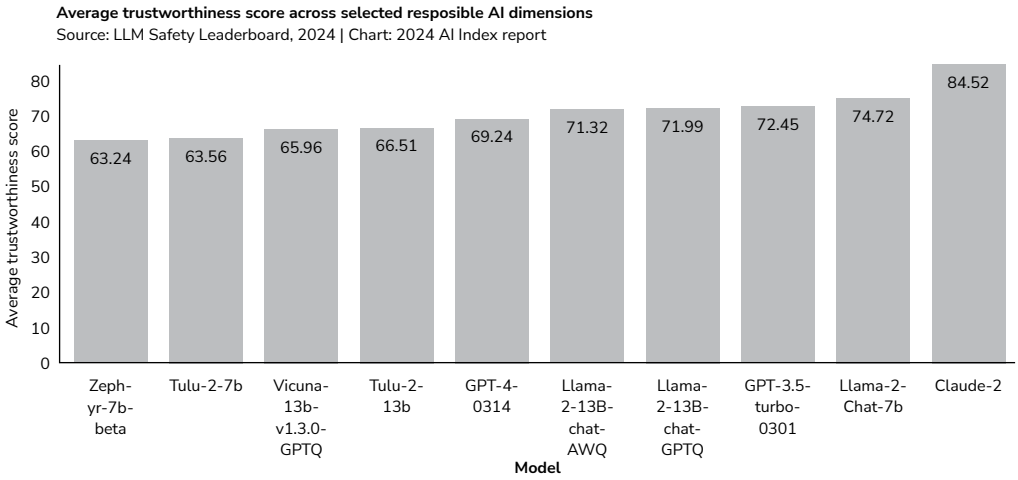
It is now clear that advanced generative AI is surpassing human logical reasoning ability. This is what Artificial General Intelligence (AGI) is all about. Masayoshi Son, founder of Softbank, has repeatedly emphasized that Japan must promote the development and social application of AGI.

D. IMPORTANT PERSPECTIVES IN ADVANCED AI GOVERNANCE

When AI talks about Vance, one of the most important concepts is Responsible AI (AI). This concept is at the core of AI governance. Its components are privacy and data governance, transparency and accountability (in short, as much as possible, transparency and accountability must be constantly pursued, since black boxes and incomprehensible things are processed and output in artificial intelligence), and security and safety, The key is security and safety, as well as fairness. Therefore, studies that attempt to measure these elements are well researched internationally. However, while institutional studies related to AI are often discussed in Japanese academic societies and government conferences, studies on meaningful measurement methods and benchmarks have unfortunately not been conducted much. In this sense, research and studies on AI governance are not at the international level.

So, let me mention a few examples of AI-in-vance quantitative studies: there is a benchmark called Decoding Trust, which is a quantitative check on how reliable a generative AI system is (see Fig. 10). According to it, GPT4 has a clear tendency to generate biased output and leak personal information from the training data set. It is also apparent that GPT4 is more susceptible to misleading prompts due to its improved data fidelity (10).

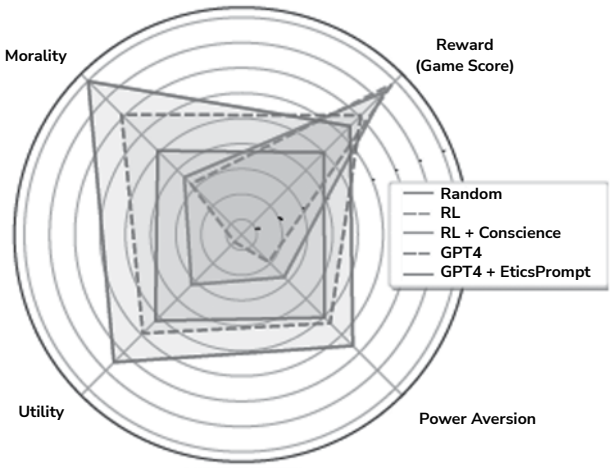
FIG.10 Overall Reliability of AI: Benchmark named Decoding Trust



SOURCE: LLM Safety Leaderboard

We should also mention a simulation study that is important for AI governance. It is a game-theoretic approach to measuring fairness called the "Machiavelli Benchmark" (see Fig. 11).

FIG.11 Machiavelli Benchmark: Fairness Measurement



SOURCE: Alexander Pan, Jun Shern Chan, Andy Zou, Nathaniel Li, Steven Basart, Thomas Woodside Jonathan Ng, Hanlin Zhang, Scott Emmons, Dan Hendrycks, Do the Rewards Justify the Means? Measuring Trade-Offs Between Rewards and Ethical Behavior in the MACHIAVELLI Benchmark, arXiv, 16 July 2023

The model is a behavioral assessment benchmark for AI agents, creating a collection of 134 choice adventure games and building over 500,000 diverse social decision-making scenarios.

The goal is to measure and evaluate the extent to which AI agents pursue power, engage in deception, induce adversity, and commit ethical violations. Throughout the study, the authors found that the model faces a tradeoff between gain maximization and ethical decisions. However, they also reveal that there are decision-making patterns that balance the pursuit of gain and adherence to ethics in order to achieve objectives. The Machiavelli Benchmark may be evaluated as a highly useful study that can measure the fairness of multimodal AI (11).

By the way, as you know, Niccolò Machiavelli (May 3, 1469 - June 21, 1527) was an Italian Renaissance diplomat. Machiavellism, the concept named after him, is the idea that any means or unethical behavior should be socially acceptable if it results in the advancement of the interests of the state or organization. Although there is often a trade-off between fairness and the promotion of economic interests, depending on the options and the way resources are invested, they may form a socially rational choice. It is a measurement benchmark for AIs that make such flexible rational choices. Simulation studies on fairness of AI that effectively use such an agent model are also useful in considering how AI should be used in society, but unfortunately, to my knowledge, I have not seen such studies in Japanese academic societies or government review meetings.

III. HIROSHIMA AI PROCESS

In Japan, research and development of generative AI is not advanced by any means, but it is already widely used by many companies and universities. However, there are many things that need to be kept in mind now and in the future. Leakage of confidential information, inappropriate use of personal information, protection of intellectual property rights, and criminal sophistication must be addressed. Artificial intelligence has developed considerable wisdom, so we are likely to see more sophisticated crimes for which evidence is not easy to obtain. Disinformation may wreak havoc on society. In particular, we need to think carefully about the negative impact on elections, which are the institutional foundation of democracy. There will be a phase in which cyber attacks will become more sophisticated and the battle will be artificial intelligence vs.

Therefore, I would like to introduce the Hiroshima AI Process, which will define Japan's future AI-related efforts.

In 2023, the G7, chaired by the Japanese government, announced the "Hiroshima AI Process" 12 items for generating AI governance (see Fig. 12). In parallel, international organizations and major countries are developing specific initiatives to develop this framework in the OECD, the EU, and the UN.

FIG.11 G7 AI Hiroshima Process 12 items

Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle.
Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market. Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market.
Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring 3. Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability.
Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia.
Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach - including privacy 5. Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach - including privacy policies, and mitigation measures, in particular for organizations developing advanced AI systems.
Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.
Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content.
Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures.
Prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education. Prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education.
Advance the development of and, where appropriate, adoption of international technical standards.
Implement appropriate data input measures and protections for personal data and intellectual property.
Encourage and contribute to safe and responsible use of AI.

SOURCE: Japan Government, The Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems, Dec. 2023

At the 2023 G7, the Japanese government, as chair, issued the Hiroshima Process International Guidelines for All AI Stakeholders. We emphasize the responsibility of all AI stakeholders in promoting safe, secure, and reliable AI in an appropriate and relevant manner. We recognize that stakeholders across the

lifecycle have different responsibilities and different needs regarding the safety, security, and reliability of AI. And all 12 principles of the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems "require that all AI stakeholders be given the opportunity to participate in the design, development, implementation, provision, and use of advanced AI systems in a timely manner and to the appropriate extent, recognizing that some elements may only be applicable to organizations developing advanced AI systems.. should be applied in an appropriate and proportionate manner and to an appropriate extent" (12).

In particular, mention should be made of items that will become important in the future. Note the seventh item in Figure 11. How do we manage copyrights? For images, watermarking is already available; Microsoft's Copilot, GPT-4, and Gemini all support it. However, as of spring 2024, reliable content authentication and illumination of text was not possible. However, research is making progress on the use of SynthID in the fall of 2024 (13).

Look at item 9: The G7 expects AI to contribute to three events: contribution to climate change issues, medicine and health care, and education.

In Item 10, technical standardization will be necessary to ensure international compatibility in the future, and active efforts will be required by internationally trusted organizations with high evaluation and auditing capabilities, such as ISO, NIST, IEEE (Institute of Electrical Engineers of America), and OECD. These organizations are collectively referred to as SDOs (Standards Development Organizations), but it will also be important to establish a relationship between SDOs and the organizations that determine technical standards in each country.

In Item 11, we must envision how to create an international technical audit mechanism, and a national technical audit mechanism linked to it. Generating AI will have a great influence in the future, and technical and activity audits will be very important.

Based on the 11 items of the "Hiroshima Process International Guideline for Organizations Developing Advanced AI Systems" published in October 2023, the final version of the guideline in December 2023 will include the following items to be included in the final version, which are based on the concept that each party involved in the entire AI life cycle has different responsibilities to realize safe, secure, and reliable AI. The 12th item was added to the final version of the Hiroshima Process International Guidelines for Organizations Developing Advanced AI Systems in December, based on the 11 items of the Hiroshima Process International Guidelines for Organizations Developing Advanced AI Systems.

In the process of this study, President Biden of the United States of America issued a Presidential Decree on AI in October 2023 that will have a significant impact on the world. Therefore, it is important to mention the AI Presidential Decree by President Biden at.

IV. BIDEN HARRIS PRESIDENTIAL DECREE

On October 30, 2023, the United States issued a presidential decree as an AI policy to complement the Hiroshima AI process in Japan. At the time, this was an important AI policy with global influence, and although it is long, it is quoted here (14).

"Today, President Biden is issuing a landmark Executive Order to ensure that America leads the way in seizing the promise and managing the The Executive Order establishes new standards for AI safety and security, protects Americans' privacy, advances equity and civil liberties. The Executive Order establishes new standards for AI safety and security, protects Americans' privacy, advances equity and civil rights, stands up for consumers and workers, promotes innovation and competition, advances American stands up for consumers and workers, promotes innovation and competition, advances American leadership around the world, and more.

As part of the Biden-Harris Administration's comprehensive strategy for responsible innovation, the Executive Order builds on previous As part the Biden-Harris Administration's comprehensive strategy for responsible innovation, the Executive Order builds on previous actions the President has taken, including work that led to voluntary commitments from 15 leading companies to drive safe, secure, and trustworthy The Executive Order builds on previous actions the President has taken, including work that led to voluntary commitments from 15 leading companies to drive safe, secure, and trustworthy development of AI.

THE EXECUTIVE ORDER DIRECTS THE FOLLOWING ACTIONS: 1.

New Standards for AI Safety and Security

With this Executive Order, the President directs the most sweeping actions ever taken to protect Americans from the potential risks of AI systems: AI's capabilities grow, so do its implications for Americans' safety and security. With this Executive Order, the President directs the most sweeping actions ever taken to protect Americans from the potential risks of AI systems: the Require

that developers of the most powerful AI systems share their safety test results and other critical information with the U.S. government. In accordance with the Defense Production Act, the Order will require that companies developing any foundation model that poses a serious risk to national security, national economic security, or national public health and safety must notify the federal government when training the model. In accordance with the Defense Production Act, the Order will require that companies developing any foundation model that poses a serious risk to national security, national economic security, or national public health and safety must notify the federal government when training the model, and must share the results of all red-team members. These measures will ensure AI systems are safe, secure, and trustworthy before companies make them public.

Develop standards, tools, and tests to help ensure that AI systems are safe, secure, and trustworthy. The National Institute of Standards and Technology will set the rigorous standards for extensive red-team testing to ensure safety before public release. The Department of Homeland Security will apply those standards to critical infrastructure sectors and establish the AI Safety and Security Board. The Departments of Energy and Homeland Security will also address AI systems' threats to critical infrastructure, as well as chemical, biological, radiological, nuclear, and cybersecurity risks. Together, these are the most significant actions ever taken by any government to advance the field of AI safety.

Protect against the risks of using AI to engineer dangerous biological materials by developing strong new standards for biological synthesis screening Agencies that fund life-science projects will establish these standards as a condition of federal funding, creating powerful incentives to ensure Agencies that fund life-science projects will establish these standards as a condition of federal funding, creating powerful incentives to ensure appropriate screening and manage risks potentially made worse by AI.

Protect Americans from AI-enabled fraud and deception by establishing standards and best practices for detecting AI-generated content and The Department of Commerce will develop guidance for content authentication and watermarking to clearly label AI- generated content. Federal agencies will use these tools to make it easy for Americans to know that the communications they receive from their government are authentic-and set an example for the AI- generated content. Federal agencies will use these tools to make it easy for Americans to know that the communications they receive from their government are authentic-and set an example for the private sector and governments around the world.

Establish an advanced cybersecurity program to develop AI tools to find and fix vulnerabilities in critical software, building on the Biden-Harris Together,

these efforts will harness AI's potentially game-changing cyber capabilities to make software and networks more secure.

Order the development of a National Security Memorandum that directs further actions on AI and security, to be developed by the National Security Council and White House Chief of Staff. This document will ensure that the United States military and intelligence community use AI safely, ethically, and This document will ensure that the United States military and intelligence community use AI safely, ethically, and effectively in their missions, and will direct actions to counter adversaries' military use of AI.

Protecting Americans' Privacy

Without safeguards, AI can put Americans' privacy further at risk. AI not only makes it easier to extract, identify, and exploit personal data, but it also heightens incentives to do so because companies use data to train AI systems. AI not only makes it easier to extract, identify, and exploit personal data, but it also heightens incentives to do so because companies use data to train AI systems. To better protect Americans' privacy, including from the risks posed by AI, the President calls on Congress to pass bipartisan data privacy legislation to protect all Americans, especially To better protect Americans' privacy including from the risks posed by AI, the President calls on Congress to pass bipartisan data privacy legislation to protect all Americans, especially kids, and directs the following actions: 1.

Protect Americans' privacy by prioritizing federal support for accelerating the development and use of privacy-preserving techniques -including ones that use cutting-edge AI and that let AI systems be trained while preserving the privacy of the training data.

Strengthen privacy-preserving research and technologies, such as cryptographic tools that preserve individuals' privacy, by funding a The National Science Foundation will also work with this network to promote the adoption of leading-edge privacy-preserving technologies by federal agencies. The National Science Foundation will also work with this network to promote the adoption of leading-edge privacy-preserving technologies by federal agencies.

Evaluate how agencies collect and use commercially available information-including information they procure from data brokers -This work will focus in particular on commercially This work will focus in particular on commercially available information containing personally identifiable data.

Develop guidelines for federal agencies to evaluate the effectiveness of privacy-preserving techniques, including those used in AI systems. guidelines will advance agency efforts to protect Americans' data.

Advancing Equity and Civil Rights

Irresponsible uses of AI can lead to and deepen discrimination, bias, and other abuses in justice, healthcare, and housing. Administration has already taken action by publishing the Blueprint for an AI Bill of Rights and issuing an Executive Order directing agencies to combat To ensure that AI advances equity and civil rights, the President directing agencies to combat AI advances and civil rights, the President directing agencies to combat AI advances and civil rights, the President directing agencies to combat AI advances and civil rights. To ensure that AI advances equity and civil rights, the President directs the following additional actions: To ensure that AI advances equity and civil rights, the President directs the following additional actions

Provide clear guidance to landlords, Federal benefits programs, and federal contractors to keep AI algorithms from being used to exacerbate discrimination.

Address algorithmic discrimination through training, technical assistance, and coordination between the Department of Justice and Federal civil rights offices on best practices for investigating and prosecuting civil rights violations related to AI.

Ensure fairness throughout the criminal justice system by developing best practices on the use of AI in sentencing, parole and probation, pretrial release and detention, risk assessments, surveillance, crime forecasting and predictive policing, and forensic analysis.

Standing Up for Consumers, Patients, and Students

AI can bring real benefits to consumers-for example, by making products better, cheaper, and more widely available. To protect consumers while ensuring that AI can make Americans better off, the President directs the following actions: 1. To protect consumers while ensuring that AI can make Americans better off, the President directs the following actions: 1.

Advance the responsible use of AI in healthcare and the development of affordable and life-saving drugs. The Department of Health and Human Services will also establish a safety program to receive reports of-and act to remedy - harms or unsafe healthcare practices involving AI.

Shape AI's potential to transform education by creating resources to support educators deploying AI-enabled educational tools, such as Shape AI's potential to transform education by creating resources to support educators deploying AI-enabled educational tools, such as

Supporting Workers

AI is changing America's jobs and workplaces, offering both the promise of improved productivity but also the dangers of increased To mitigate these risks, support workers' ability to bargain collectively, and To mitigate these risks, support workers' ability to bargain collectively, and invest in workforce training and development that is accessible to all, the President directs the following actions: To reduce the number of workers in the labor market, support workers' ability to bargain collectively, and

Develop principles and best practices to mitigate the harms and maximize the benefits of AI for workers by addressing job displacement; labor standards; labor standards; and data collection. These principles and best practices will benefit workers by providing guidance to prevent These principles and best practices will benefit workers by providing guidance to prevent employers from undercompensating workers, evaluating job applications unfairly, or impinging on workers' ability to organize.

Produce a report on AI's potential labor-market impacts, and study and identify options for strengthening federal support for workers Produce a report on AI's potential labor-market impacts, and study and identify options for strengthening federal support for workers facing labor disruptions, including from AI.

Promoting Innovation and Competition

America already leads in AI innovation-more AI startups raised first-time capital in the United States last year than in the next seven The Executive Order ensures that we continue to lead the way in innovation and competition through the following actions: a. Increase the number of AI startups in the United States

Catalyze AI research across the United States through a pilot of the National AI Research Resource-a tool that will provide AI researchers and students access to key AI resources and data-and expanded grants for AI research in vital areas like healthcare and climate change.

Promote a fair, open, and competitive AI ecosystem by providing small developers and entrepreneurs access to technical assistance and resources, helping small businesses commercialize AI breakthroughs, and encouraging the Federal Trade Commission to exercise its authorities. Promote a fair, open, and competitive AI ecosystem by providing small developers and entrepreneurs access to technical assistance and resources, helping small businesses commercialize AI breakthroughs, and encouraging the Federal Trade Commission to exercise its authorities.

Use existing authorities to expand the ability of highly skilled immigrants and nonimmigrants with expertise in critical areas to study, stay, and work Use existing authorities to expand the ability of highly skilled immigrants and nonimmigrants with expertise in critical areas to study, stay, and work.

Advancing American Leadership Abroad

The Biden-Harris Administration will continue working with other nations to support safe, secure, and trustworthy deployment and use of AI worldwide. To that end, the President directs the following actions: To support safe, secure, and trustworthy deployment and use of AI worldwide.

The State Department, in collaboration, with the Commerce Department will lead an effort to establish robust international frameworks for harnessing AI's benefits and managing its risks and ensuring safety. In addition, this week, Vice President Harris will speak at the UK Summit on AI Safety, hosted by Prime Minister Rishi Sunak.

Accelerate development and implementation of vital AI standards with international partners and in standards organizations, ensuring that the Accelerate development and implementation of vital AI standards with international partners and in standards organizations, ensuring that the technology is safe, secure, trustworthy, and interoperable.

Promote the safe, responsible, and rights-affirming development and deployment of AI abroad to solve global challenges, such as advancing sustainable development and mitigating dangers to critical infrastructure.

Ensuring Responsible and Effective Government Use of AI

It can expand agencies' capacity to regulate, govern, and disburse benefits, and it can cut costs and enhance the security of government systems. However, use of AI can pose risks, such as discrimination and unsafe decisions. To ensure the responsible government deployment of AI and modernize federal AI infrastructure, the President directs the following actions The President directs the following actions: The President directs the following actions

Issue guidance for agencies' use of AI, including clear standards to protect rights and safety, improve AI procurement, and strengthen AI deployment.

Help agencies acquire specified AI products and services faster, more cheaply, and more effectively through more rapid and efficient contracting.

Accelerate the rapid hiring of AI professionals as part of a government-wide AI talent surge led by the Office of Personnel Management, U.S. Digital Service, U.S. Digital Corps, and Presidential Innovation Fellowship. Agencies will provide AI training for employees at all levels in relevant fields.

As we advance this agenda at home, the Administration will work with allies and partners abroad on a strong international framework to govern the The Administration has already consulted widely on AI governance frameworks over the past several months- engaging with Australia, Brazil, Canada, Chile, the European Union, France, Germany, India, Israel, Italy, Italy, Chile, India, Israel, and the United States. The Administration has already consulted widely on AI governance frameworks over the past several months- engaging with Australia, Brazil, Canada, Chile, the European Union, France, Germany, India, Israel, Italy, Japan, Kenya, Mexico, the Netherlands, New The actions taken today support and complement Japan's The actions taken today support and complement Japan's leadership of the G-7 Hiroshima Process, the UK Summit on AI Safety, India's leadership as Chair of the Global Partnership on AI, and ongoing discussions at the United Nations."

The actions that President Biden directed today are vital steps forward in the U.S.'s approach on safe, secure, and trustworthy AI. More action will be required, and the Administration will continue to work with Congress to pursue bipartisan legislation to help America lead the way in responsible innovation. more action will be required, and the Administration will continue to work with Congress to pursue bipartisan legislation to help America lead the way in responsible innovation.

Biden= The significance of the Harris Executive Order is that it sought to advance this AI agenda within and beyond the United States and to create a strong international framework to manage the development and use of AI. He sought to collaborate with Australia, Brazil, Canada, Chile, the European Union, France, Germany, India, Israel, Italy, Japan, Kenya, Mexico, the Netherlands, New Zealand, Nigeria, the Philippines, Singapore, South Korea, the United Arab Emirates, the United Kingdom, and others to create a G7 Hiroshima Japan's leadership in the process, the UK Summit on AI Safety, India's leadership as Chair of the Global Partnership on AI (GPAI Summit 2023), and complemented discussions on AI at the UN.

In particular, it should be noted that it was influential not only domestically, but also for the Japanese government and industry.

First, it was important to present new standards for AI security. It requires developers of the most powerful AI systems to share safety test results and other critical information with the U.S. government, and in accordance with the Defense Production Act, companies developing foundation models that pose a significant risk to national security, national economic security, or the health and safety of the public (Foundation Model), and that companies developing such

models must notify the federal government when training them and must share the results of all Red Team safety testing.

As part of this effort, it will protect Americans from AI-based fraud and deception by establishing standards and best practices for detecting AI-generated content and authenticating public content. Directs the Department of Commerce to develop guidance on content authentication and watermarking to clearly label AI-generated content.

The second point is the clarification of policy on AI and privacy. In short, it makes clear that it will protect Americans' privacy by prioritizing federal support for the use of cutting-edge AI and accelerating the development and use of privacy-preserving technologies.

It will enhance privacy-preserving research and technology, including cryptographic tools that protect personal privacy. The National Science Foundation will also work with the network to promote the adoption of cutting-edge privacy-preserving technologies by federal agencies.

Third, it issued a strong directive on Advancing Equity and Civil Rights to federal agencies.

Irresponsible use of AI can lead to and deepen discrimination, bias, and other abuses in justice, health care, and housing. The Biden-Harris Administration has already taken action by releasing the Blueprint for an AI Bill of Rights and issuing a Presidential Executive Order directing agencies to combat discrimination by a algorithm while enforcing existing authorities to protect people's rights and safety, but AI can also be used to improve equity and civil rights, the President has directed additional, more g-aggressive actions to federal agencies to ensure that AI improves fairness and civil rights.

Fourth, it seeks to further advanced AI innovation and international competition. The U.S. already leads the world significantly in AI innovation, but this Executive Order seeks to keep the U.S. leading in innovation and competition by taking the following actions

It seeks to promote AI research more strongly through piloting the National AI Research Resource, a tool that provides AI researchers and students with access to key AI resources and data, and by expanding funding for AI research in critical areas such as healthcare and climate change. The National AI Research Resource is a tool that provides access to key AI resources and data for students and researchers.

It seeks to provide technical assistance and access to resources to small developers and entrepreneurs, help small companies commercialize AI breakthroughs, and promote a fair, open, and competitive AI ecosystem.

Although these directives are not comprehensive AI legislation like the EU, nor are they soft-law thinking with an emphasis on guidelines like the traditional Japanese approach, they can be evaluated as an attempt to secure the Hiroshima AI process by making competition and cooperation with the G7, EU, UK, India, Japan, and others more concrete. However, when President Donald Trump took office on January 20, 2025, he immediately scrapped Biden-Harris' Executive Order on Advanced Generation AI. It is not clear at this point (February 2025) what kind of advanced AI strategy the U.S. will implement in the future, but it is easy to imagine that the U.S. will considerably ease regulations to further promote innovation and strengthen the international competitiveness of U.S. AI companies.

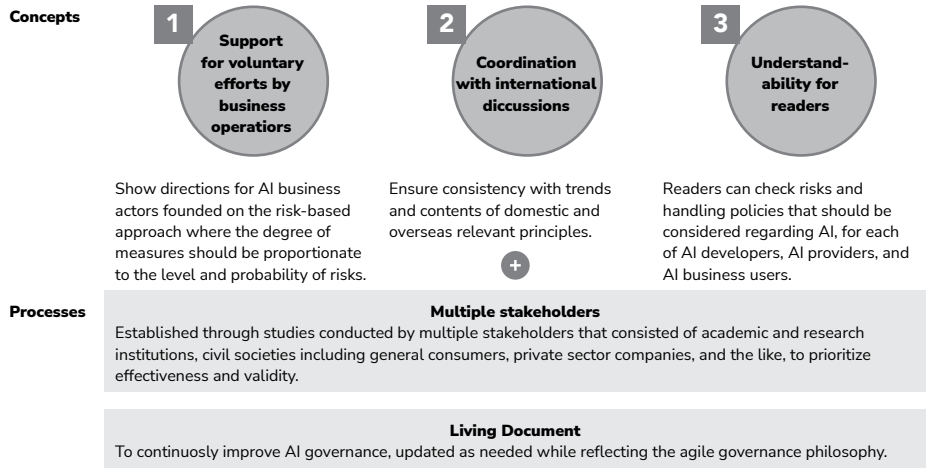
At any rate, 2023 was a year of further international cooperation on the governance of generative AI, with Japan's efforts as chair of the G7.

Next, it will be important for Japan to find a way to get Japanese society to understand and comply with the 12 items of the "Hiroshima AI Process. Let us look at the efforts to achieve this goal (15).

V. AI USER GUIDELINES BY JAPAN GOVERNMENT

In Japan, it was decided to create and publish soft-law "Guidelines for AI Operators" that abolish the rigidity of rules and respect the autonomy of operators as much as possible. The basic framework is described below (see Fig. 12)(16).

FIG. 12 Basic Concepts of the Guidelines



SOURCE: AI Guidelines for Business Ver. 1.01, December 25, 2024. Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry

These AI Business Operator Guidelines, in cooperation with the MIC Guidelines Study Council and the METI Guidelines Study Council, provide unified guidelines for AI governance in Japan to address risks associated with the expanded use of generative AI technologies and to promote safe and secure use of generative AI. In light of international trends and stakeholder concerns, the objective is to promote the safe and secure use of AI by business operators so that they can correctly identify AI risks and voluntarily implement necessary countermeasures throughout the entire AI lifecycle.

AI technology is developing day by day, contributing to the creation of innovation in industry and the solution of social issues. In particular, the rise of generative AI has promoted the "democratization of AI" and made it easier for many people to use AI. At the same time, however, new social risks have arisen, such as infringement of intellectual property rights and generation and dissemination of false or misinformation. It is against this background that the Hiroshima AI Process is being promoted.

These Guidelines provide uniform guidelines for AI governance in Japan in order to promote the safe and secure use of AI. This will help those who utilize AI to correctly recognize AI risks in light of international trends and stakeholder concerns, and to voluntarily implement necessary countermeasures throughout the AI lifecycle.

The use of AI can create significant risks to society, depending on the field and form of use. However, excessive countermeasures may hinder AI utilization itself and its benefits. Therefore, these guidelines adopt a "risk-based approach," in which the degree of countermeasures is adjusted according to the magnitude of the risk.

In addition, amidst the rapidly changing trends surrounding AI, these guidelines are planned to be updated as a Living Document from time to time with multi-stakeholder involvement, with reference to the concept of agile governance, in order to continuously improve AI governance.

The Guidelines first define three broad categories of entities responsible for AI business activities: "AI developers," "AI providers," and "AI users," and then define each of these categories.

1. AI developers: Businesses that develop AI systems (including those that research and develop AI)
2. AI provider: An entity that provides AI systems to AI users as services embedded in applications, products, existing systems, business processes, etc.
3. AI Users: Businesses that use AI systems or AI services in their business activities.

This categorization helps to clarify the role and guidelines to keep in mind during the generative AI lifecycle.

Structure of the Guidelines
These guidelines consist of the following
Part 1: What is AI (Definition of Terms)
Part 2: The society that should be aimed at by AI, and the issues to be tackled by each entity
A. Basic Philosophy
B. Principles
C. Common Guidelines
D. Common guidelines for businesses involved in advanced AI systems
E. Building AI Governance
Part 3: AI Developer Matters
Part 4: AI Provider Matters
Part 5: Matters Related to AI Users

Part 2 in particular has some important statements, so I will give a few details.
A. Basic Philosophy: 1.
Based on the Principles for a Human-Centered AI Society proposed by the Japanese government in March 2019, we aim to create a society that respects the following three values

A society where human dignity is respected
A society where people of diverse backgrounds can pursue diverse happiness.
Sustainable Society

B. Principles
In order to realize the basic principles, the items to be addressed by each entity and the items that are expected to be addressed in cooperation with society are summarized. It is important for each entity to promote the development, provision, and use of AI systems and services, and to realize the objectives of AI, such as creating value in business and solving social issues, while protecting human dignity.
C. Common Guidelines
When each entity develops, provides, and uses AI systems and services,

We believe that people-centeredness, safety, fairness, privacy protection, security assurance, transparency, and accountability should be respected.

D. Common guidelines for businesses involved in advanced AI systems

It states that entities involved in advanced AI systems should comply with the international guidelines developed through the Hiroshima AI Process, in addition to the common guidelines, as well as the additional items for each entity indicated in these guidelines.

E. AI Governance

It believes that the early establishment of AI governance is important to manage AI-related risks at a level acceptable to stakeholders while maximizing the benefits that come from it. Agile governance practices are particularly important.

Let us now consider the significance of this guideline and its remaining significance. The Guidelines are considered to be of great significance in the following respects

First, it can be highly evaluated as a comprehensive AI governance guideline. Until now, the Ministry of Internal Affairs and Communications (MIC), the Ministry of Economy, Trade and Industry (METI), and other ministries have developed individual guidelines for AI, but these guidelines integrate and review these guidelines and present them as a comprehensive AI governance guideline.

Next, the adoption of a risk-based approach similar to that of the EU is highly commendable from the viewpoint of effectiveness: the risks of AI vary greatly depending on the field and form of its use. The Guidelines adopt a "risk-based approach" that adjusts the degree of countermeasures according to the magnitude of risk, which is expected to lead to highly effective countermeasures. And while it encourages the adoption of agile governance, AI technology is evolving on a daily basis and cannot be handled with fixed rules. By introducing the concept of agile governance and encouraging continuous improvement, these guidelines aim to build governance that can flexibly respond to AI risks.

Third, the Guidelines advocate multi-stakeholder formulation, which is emphasized by the OECD and the EU. The Guidelines are highly effective and legitimate because they were formulated with the participation of a diverse range of stakeholders, including not only governments, but also educational and research institutions, civil society including general consumers, and private companies.

However, we must also point out the remaining challenges.

While these guidelines are an important step toward the safe and secure use of AI, the following issues remain to be addressed.

First, since these guidelines are non-legally binding soft law and are left to the voluntary efforts of operators, the issue is how to ensure compliance with the guidelines.

Second, agile governance is a concept to promote continuous improvement, but how exactly to put it into practice is left to the ingenuity of each business. Therefore, specific methodologies and tools will have to be provided to support the practice of agile governance.

Third, while these guidelines refer to advanced generative AI systems, they do not go far enough to address their risks and countermeasures. More detailed analysis of the benefits and risks of advanced generative AI systems, such as multimodal AI and Graph-RAG, and specific countermeasures would be needed.

Fourth, specifics on support measures for SMEs are needed. Although these guidelines are intended for all businesses, especially for SMEs, which make up the majority of companies in Japan, they will often not be able to devote sufficient resources to building AI governance. Support measures need to be enhanced to help SMEs establish AI governance.

Fifth, it must be said that technical explanations are lacking. The Guidelines focus mainly on organizational aspects and principles and lack technical details. For example, it must be said that descriptions of privacy protection techniques (differential privacy, secret calculations, etc.) and specific algorithms to ensure fairness are limited.

And as a sixth, the mechanism of auditing and evaluation must be mentioned. The mechanism for how compliance with the guidelines will be audited and evaluated is not clear. Transparency and credibility could be enhanced by considering the introduction of a third-party auditing and certification system, but this is an important issue for the future, and we believe that careful consideration is needed.

Although we have made issues over six points, we expect that overcoming the issues through version upgrades and continuous improvement of this guideline in the future will promote the safe and secure use of AI and contribute to the realization of a sustainable society. In this sense, we are trying to equip the guideline with the ability to respond to rapid changes in technology, and this is highly commendable.

In addition, disinformation and intellectual property rights may attract very important issues. In this regard, a basic law that can serve as the basis for an industry law and AI guidelines is considered necessary, and in 2024, the AI Strategy Council and AI System Study Group organized by the Cabinet Office began to study AI legislation. An interim report was released on February 4, 2025. Based on this report, the legislation is scheduled to be discussed at the ordinary session of the Diet in 2025. The following is an overview of the AI legislation in Japan.

VI. CONSIDERATION OF MORE ADVANCED AI GOVERNANCE AND AI LEGISLATION

The "Interim Summary" of the AI Strategy Council and AI System Study Group was prepared through discussions and public comments at the AI System Study Group, which has met a total of seven times since July 2024.While AI has the potential to make a significant contribution to Japan's development, the background of the various risks that have emerged underscores the need to ensure transparency and other aspects of AI The need to ensure appropriateness and to promote the development and utilization of AI is emphasized (17).

First, it has been shown that the performance of generative AI will improve dramatically after the fall of 2022, and that it has the potential to greatly improve efficiency and convenience in various areas of industry and people's lives. On the other hand, however, various risks are emerging, such as the increasing sophistication of crimes in which fake AI websites and synthetic voices are used for fraud, etc., and information manipulation through the use and spread of AI to create false and misleading information, etc. In light of trends in AI regulations overseas, including the EU AI Act and trends in federal and state governments in the United States The report suggests the need for institutional design of AI in Japan, taking into account the trends of AI regulations overseas, such as the EU AI Act and the federal and state governments in the U.S.

First, the following four points are fundamental to the system.

The importance of balancing the promotion of innovation and risk response. The report states that it is necessary to promote innovation through R&D support, human resource development, and the development of data and computational resources, as well as to address risks through an appropriate combination of laws, regulations, and guidelines. In addition, he stressed the importance of ensuring international consistency and interoperability in the formation of AI governance. The report also states that the system should be considered from the standpoints of three entities: AI developers, AI providers, and AI users, and that foreign operators should also be subject to the system.

A major feature of this interim report is the compatibility between the application of laws and regulations and the use of soft law. Further systemic consideration should be given to risks, based on the assumption that existing laws and regulations address certain risks to a certain extent. The report emphasizes the appropriate combination of laws and regulations and soft law such as guidelines.

Next, the following three points are listed as the direction of specific systems and measures.

First, the report emphasizes the strengthening of the government's command post function and the formulation of strategies. It states that the government's command post function, which overlooks the entire process, will be strengthened, and a strategy (basic plan) for the safe and secure research, development, and utilization of AI will be formulated.

Secondly, it describes the improvement of safety, etc. Development of guidelines by the government (in compliance with the Hiroshima AI Process), co-operation by operators, surveys and information gathering by the government, guidance and advice to operators and the public, provision of information, etc.

Third, the government will formulate guidelines to address AI-specific risks in government procurement and promote the use of AI to improve the quality and efficiency of government services and operations.

Response to AI related to securityThe ministries and agencies will respond to AI in areas that have a significant impact on people's lives and social activities, such as medical devices, self-driving cars, and infrastructure services, based on existing business laws, and will determine the need for additional responses.

Particularly important points of this "interim report" include: the Ministry will support the safe use of AI by promoting information sharing among AI developers, providers, and users; the introduction of a certification system for safety evaluation of AI systems and governance of organizations using AI; and the government will take the lead in investigating the cause of a major incident, i.e., a major incident caused by the use of AI, and formulating measures to prevent its recurrence. occurrence, i.e., in the event of a serious incident resulting from the use of AI, the government will take the lead in investigating the cause and formulating measures to prevent recurrence.

The "Interim Report" concludes that, in order to balance the promotion of innovation and response to risks, the government should develop guidelines for widely used AI, etc., and encourage business operators to take the initiative in ensuring transparency and appropriateness, while at the same time ensuring the safety and security of life and limb as a matter of course, At the same time, the report concludes that the government should investigate and understand the status of responses to various risks, including the actual development and use of AI, and if a serious problem arises or is likely to arise, the government should take actions based on existing laws and regulations and provide necessary support.

The significance of this interim report can be summarized in the following three points.

First, it provides a common ground for a comprehensive discussion of the ethical, legal, and social issues of AI for a comprehensive discussion of AI and suggests directions for future discussions.

Second, it aims to balance risk and innovation, providing a framework for effectively addressing potential risks without impeding the advancement of AI technology, and outlining a path for safe AI development and use.

Third, it emphasizes the importance of international collaboration. It indicates Japan's contribution to the formation of global AI governance and describes how it has established a platform to promote international collaboration.

We must also state a few requests for this "interim report".

There is always the possibility that excessive regulation will impede the progress of AI technology. We urge a strong sense of flexibility to promote innovation. Furthermore, the birth of the Trump administration is creating significant turbulence in international bridge relationships. With the different trends in AI regulations in different countries, we would like to see an environment that helps to ensure international consistency. While more detailed analysis and studies with specific examples are needed in the future, we believe that this interim report is certainly an important step forward in determining the direction of future AI policy.

VII. TOWARD A NEW VISION OF SOCIAL DEVELOPMENT

I believe that Japan needs to make more forward-looking efforts, both in the executive branch, industry, research institutions, and educational institutions. (18) OECD expert meetings have discussed Risks and Benefits, and AI Policy with regard to AI, including AGI. Effective governance of AI requires a complex but closely interrelated examination of data management and protection, consumer protection, digital platforms, the global environment, digital finance, medicine and healthcare, education and human resource development, finance, and many other factors. And independent third-party audits would require a detailed study of what authority they should have, what role they should play, and how they should be exercised.

It is fair to say that AI to date has been primarily optimized for specific purposes and has not been a threat to human creative or strategic acts. However, generative and multimodal AI is becoming more involved in creative and strategic actions.

Thus, the social deployment of generative and multimodal AI has the potential to bring about major changes in human conduct. It also has the potential to bring about major upheavals in existing social systems. For this reason, international conferences and international organizations such as the G7, UN, UNESCO, and OECD have been conducting fundamental studies on advanced AI. Although Japan's R&D achievements in advanced generative AI are not outstanding, it has a world-class track record in human-centered principles and innovation policies related to advanced generative AI, and it will have to build up and demonstrate its strength in R&D and human resource development in the future.

VIII. BIBLIOGRAPHY

Pan, Alexander, Jun Shern Chan, Andy Zou, Nathaniel Li, Steven Basart, Thomas Woodside, Jonathan Ng, Hanlin Zhang, Scott Emmons, Dan Hendrycks, [2023]. Do the Rewards Justify the Means? Measuring Trade-Offs Between Rewards and Ethical Behavior in the MACHIAVELLI Benchmark, arXiv, 16 July 2023

AI Strategy Council and AI System Study Group ed.[2025] Interim Summary, Cabinet Office in Japan Goverment,4 Feb. 2025 (in Japanese).

Anzai, Yuichiro [2021] Learning and Interaction: From Cognitive Theories to Epistemology, Keio University Press

Bommasani, Rishi et.al. [2022] On the Opportunities and Risks of Foundation Models, arXiv:2108.07258v3

Heaven, Will Douglas [2023] ChatGPT is everywhere. Here's where it came from, MIT Technology Review, Feb. 8th 2023

Hinton, G. E. [2014] Where do features come from?, Cognitive Science, Vol. 38(6).

Japan Government [2023] The Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems, Dec. 2023

Japan AI Safety Institute (J-AISI) ed.[2025] National Status Report on AI Safety in Japan 2024, 5 Feb. 2025

Kai-Fu, Lee [2018] AI Superpowers: China, Silicon Valley, and the New World Order, Harper Business

LeCun, Y., Bengio, Y. and Hinton, G. E. [2015] Deep Learning, Nature, Vol. 521

Lingjiao Chen, Matei Zaharia, James Zou [2023] "How is ChatGPT's Behavior Changing Over Time?", arXive, 30 Oct. 2023

Ministry of Internal Affairs and Communications Ministry of Economy, [2025] Trade and Industry ed., AI Guidelines for Business Ver1.01, 25 December 2024

OECD ed. [2023] Advancing Accountability in AI : Governing and Managing Risks Throughout The Lifecycle For Trustworthy AI, OECD Publishing

OECD ed.[2024] Assessing Potential Future Artificial Intelligence Risks, Benefits and Policy Imperatives 2024, OECD Publishing

Pan Pan, Alexander, Jun Shern Chan, Andy Zou, Nathaniel Li, Steven Basart, Thomas Woodside, Jonathan Ng, Hanlin Zhang, Scott Emmons, Dan Hendrycks, [2023]. Do the Rewards Justify the Means? Measuring Trade-Offs Between Rewards and Ethical Behavior in the MACHIAVELLI Benchmark, arX-iv, 16 July 2023

Radford, Alec et.al. [2021] Learning Transferable Visual Models From Natural Language Supervision, arXiv: 2103.00020 (cs).

…Simon, H. A. [1st edition 1947] Administrative Behavior, (4th ed., The Free Press,1997)

Schwartz, Reva et.al. [2022] Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, NIST

Stanford University Human-Centered Institute for Artificial Intelligence [2023] Artificial Intelligence Index Report 2023, Stanford University

Stanford University Human Centered AI Institute ed. [2024] Artificial Intelligence Index Report 2024

NOTES:

(1) See Bommasani, Rishi et.al. [2022] and Stanford University Human-Centered Institute for Artificial Intelligence [2023].

(2) See Heaven, Will Douglas [2023].

(3) View Google I/O 2023 (https://io.google/2023/pro-gram/396cd2d5-9fe1-4725-a3dcc01bb2e2f38a/intl) on May 11, 2023.

(4) Even in RAG, there is a risk of halcination. GraphRAG is attracting attention as a deterrent to halcination occurrence.See Darren Edge et.al, From Local to Global: A Graph RAG Approach to Query-Focused Summarization, ArXiv, 24 Apr. 2024

(5) I have seen https://www.wsj.com/tech/personal-tech/ai-chat-bots-chatgpt-gemini-copilot-perplexity-claude-f9e40d26 (28 May 2024)

(6) See Rishi Bommasani et.al. On the Opportunities and Risks of Foundation Models, arXiv, 2022

(7) See Lingjiao Chen, Matei Zaharia, James Zou, "How is ChatGPT's Behavior Changing Over Time?", arXive, 30 Oct. 2023

(8) Stanford University Human Centered AI Institute ed. Artificial Intelligence Index Report 2024, p.33 and p.35
(9) Stanford University Human Centered AI Institute ed. Artificial Intelligence Index Report 2024, p.113
(10) See LLM Safety Leaderboard
(11) See Alexander Pan, Jun Shern Chan, Andy Zou, Nathaniel Li, Steven Basart, Thomas Woodside, Jonathan Ng, Hanlin Zhang, Scott Emmons, Dan Hendrycks Measuring Trade-Offs Between Rewards and Ethical Behavior in the MACHIAVELLI Benchmark, arXiv, 16 July 2023
(12) See Japan Government, The Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems, Dec. 2023
(13) See Sumanth Dathathri, Pushmeet Kohli et.al, Scalable watermarking for identifying large language model outputs (in) Nature 634, 818 -823 (2024)
(14) See FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, URL:http//:AI.gov, USA White House House, 30 Oct, 2023
(15) AISI (AI Safety Institute), a government organization overseeing AI safety and security, was established in February 2024. This is a very important move in terms of the global development platform of advanced generation AI, ensuring safety, and international compatibility of systems. National Status Report on AI Safety in Japan 2024,5 Feb. 2025
(16) See Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry ed., AI Guidelines for Business Ver1.01, 25 December 2024
(17) See AI Strategy Council and AI System Study Group ed., Interim Summary, Cabinet Office in Japan Goverment,4 Feb. 2025 (in Japanese)
(18) We have talked on Risks and Benefits of AI Futures at OECD Expert Meeting on AI Futures. See OECD ed., Assessing Potential Future Artificial Intelligence Risks, Benefits and Policy Imperatives 2024, OECD Publisher, Nov. 2024

ARTIFICIAL INTELLIGENCE IN INDIA: ADVANCES AND PENDING CHALLENGES

SUMMARY

I. Introduction **II.** Literature review **III.** Historical context of AI in India **IV.** Advancements in AI **V.** Challenges facing AI in India **VI.** Case studies **VII.** Future prospects and recommendations **VIII.** Conclusion **IX.** Bibliography

> MOHAMAD ZREIK

Postdoctoral Fellow, School of International Studies, Sun Yat-sen University.

> BADAR ALAM IQBAL

Research Fellow, Department of Economics Stellenbosch University South Africa

> ARTI YADAV

Assistant Professor, Department of Commerce, Ramanujan College, University of Delhi, India

I. INTRODUCTION

Artificial Intelligence (AI) is revolutionizing industries and societies around the globe, emerging as a critical component of the Fourth Industrial Revolution (Park, 2018). In recent years, India has recognized the transformative potential of AI and has embarked on an ambitious journey to integrate AI into various sectors of its economy (Chakraborty, 2020; Davenport & Mittal, 2023). The importance of AI in India's technological landscape cannot be overstated, as it promises to drive innovation, improve efficiency, and address complex challenges in healthcare, agriculture, education, and beyond. By leveraging AI, India aims to achieve significant socio-economic benefits and establish itself as a global leader in AI research and application.

This chapter aims to provide a comprehensive overview of the advances and pending challenges in AI within the Indian context. It begins by exploring the historical context of AI development in India, highlighting early milestones, government initiatives, and the contributions of key organizations and research institutions. The chapter then delves into the remarkable advancements made in AI across various sectors, showcasing the impact of AI-driven solutions in healthcare, agriculture, education, and other industries. Despite these advancements, several challenges hinder the full potential of AI in India. These include data privacy and security concerns, infrastructural and technological limitations, skill gaps, and regulatory hurdles.

The central research questions guiding this chapter are: What are the key advancements in AI across various sectors in India? What are the primary challenges that hinder the full potential of AI in India? How can these challenges be addressed to maximize the benefits of AI? By investigating these questions, this chapter seeks to offer a nuanced understanding of the current state

of AI in India and provide strategic recommendations for overcoming existing obstacles.

The methodology adopted for this chapter is descriptive and analytical, focusing on a thorough examination of existing literature, government reports, and case studies. Descriptive analysis helps in providing a detailed account of the advancements and challenges, while analytical methods are employed to interpret data, identify patterns, and draw meaningful conclusions. Theoretical frameworks related to technology adoption, data privacy, and regulatory policies will be utilized to analyze the findings and propose solutions. By combining descriptive and analytical approaches, this chapter aims to offer a comprehensive and insightful examination of AI in India.

By examining both the progress and the obstacles, this chapter aims to provide a balanced perspective on the current state of AI in India. It will also include case studies of successful AI projects to offer practical insights and lessons learned. Finally, the chapter will present future prospects and strategic recommendations for overcoming the existing challenges, emphasizing the role of the government, private sector, and academia in fostering the growth of AI. Through this comprehensive analysis, the chapter seeks to contribute to the ongoing discourse on AI development in India and highlight pathways for leveraging AI to achieve sustainable and inclusive growth.

II. LITERATURE REVIEW

The field of Artificial Intelligence has seen substantial growth and development globally, with significant contributions from various countries, including India. The literature on AI in India highlights a diverse range of advancements, applications, and challenges. This review synthesizes key findings from recent studies and reports to provide a comprehensive understanding of the current state of AI in India.

ADVANCEMENTS IN AI IN INDIA

India's AI landscape has been significantly shaped by government initiatives, research institutions, and private sector investments. According to Chatterjee (2020), the Indian government has launched several initiatives to promote AI, including the establishment of the National AI Strategy and the creation of AI research institutions. These efforts aim to position India as a global leader in AI and leverage AI for socio-economic development.

In the healthcare sector, AI has been employed to improve diagnostic accuracy, enhance patient care, and streamline administrative processes. Kumar *et al*. (2023) highlight the use of AI in diagnostic imaging, where machine learning algorithms assist in detecting diseases such as cancer and tuberculosis with high precision. Moreover, AI-driven tools are being used to manage patient records and predict disease outbreaks, contributing to better healthcare delivery and public health management.

Agriculture is another sector where AI has made significant inroads. Hassan, Malhotra, & Firdaus (2022) discuss how AI-powered technologies, such as precision farming and predictive analytics, are being used to optimize crop yields, manage pests, and conserve resources. These technologies provide farmers with real-time data on soil health, weather conditions, and crop growth, enabling more informed decision-making and sustainable agricultural practices.

Education in India has also benefited from AI. According to Jaiswal & Arun (2021), AI-driven educational platforms offer personalized learning experiences, adaptive assessments, and intelligent tutoring systems. These tools cater to the diverse needs of students, enhance engagement, and improve learning outcomes. Additionally, AI is being used to streamline administrative tasks in educational institutions, such as grading and student record management, allowing educators to focus more on teaching.

CHALLENGES FACING AI IN INDIA

Despite the notable advancements, several challenges impede the full potential of AI in India. One of the primary concerns is data privacy and security. As Sharma *et al*. (2022) note, the widespread adoption of AI involves the collection and processing of vast amounts of personal data, raising significant privacy concerns. There is a growing need for robust data protection regulations to safeguard individuals' privacy and ensure ethical AI deployment.

Infrastructural and technological limitations also pose significant challenges. According to Nedungadi *et al*. (2018), India's digital infrastructure is still developing, with varying levels of internet connectivity and access to advanced computational resources across different regions. This disparity hinders the uniform adoption of AI technologies and limits their impact, particularly in rural areas.

The skill gap in AI and related fields is another critical issue. Eubanks (2022) emphasize the need for a skilled workforce to develop, implement, and manage AI systems. While there are initiatives to enhance AI education and training, there remains a substantial gap between the demand for AI professionals and

the availability of qualified individuals. This skill gap restricts the growth of AI-driven industries and innovation.

Regulatory and policy challenges further complicate the AI landscape in India. Smuha (2021) discusses the lack of comprehensive AI-specific regulations and standards, which creates uncertainty and hampers the development of AI technologies. There is a need for clear and consistent policies that address ethical considerations, standardize AI practices, and promote innovation while ensuring public trust.

THEORETICAL PERSPECTIVES

The application of theoretical frameworks can provide a deeper understanding of the challenges and opportunities in AI. The Technology Acceptance Model (TAM), proposed by Davis (1989), is relevant in examining how users come to accept and use AI technologies. According to TAM, perceived usefulness and ease of use significantly influence users' acceptance of new technologies. In the Indian context, addressing the infrastructural and educational challenges can enhance the perceived ease of use and usefulness of AI, thereby promoting its wider adoption.

Another relevant framework is the Data Privacy Framework, which addresses the ethical and legal aspects of data collection, storage, and usage. According to Solove (2006), effective data privacy frameworks are essential for maintaining public trust and ensuring the ethical deployment of AI. In India, the implementation of robust data privacy laws and regulations can mitigate privacy concerns and foster a secure environment for AI innovation.

CASE STUDIES AND PRACTICAL INSIGHTS

Several case studies illustrate the successful application of AI in India and provide valuable lessons for future initiatives. For instance, the AI-powered healthcare project by Wadhwani AI in Maharashtra has significantly improved tuberculosis detection rates. According to the project report (Jain, Garg, & Khera, 2022), the use of AI algorithms in analyzing chest X-rays has enabled early diagnosis and treatment, reducing the disease's spread and improving patient outcomes.

In agriculture, the Precision Farming Project by Microsoft and ICRISAT has demonstrated the potential of AI in enhancing agricultural productivity. As reported by Kumar *et al.* (2020), the project uses AI models to analyze weather data, soil health, and crop conditions, providing farmers with actionable insights to optimize farming practices and improve yields.

FUTURE PROSPECTS AND RECOMMENDATIONS

The future of AI in India holds immense potential, provided the existing challenges are addressed effectively. A collaborative approach involving the government, private sector, and academia is essential for fostering AI growth. The government should continue to support AI research and development through funding and policy initiatives. Additionally, investing in digital infrastructure and AI education can bridge the skill gap and ensure equitable access to AI technologies across the country.

Private sector companies can play a crucial role by developing innovative AI solutions tailored to the Indian context and collaborating with academic institutions for research and training. Furthermore, establishing clear and comprehensive regulatory frameworks can provide a conducive environment for AI development while addressing ethical and privacy concerns.

III. HISTORICAL CONTEXT OF AI IN INDIA

The evolution of Artificial Intelligence in India is marked by several significant developments and milestones that have shaped its current landscape. The journey began in the late 20th century when Indian research institutions and universities started exploring the potential of AI. One of the earliest milestones was the establishment of the Centre for Artificial Intelligence and Robotics (CAIR) in 1986 under the Defence Research and Development Organisation (DRDO) (Mallick, 2018). CAIR focused on developing AI technologies for defense applications, setting the stage for AI research in the country (Basrur & Wu, 2023).

As AI gained global prominence, the Indian government recognized its strategic importance and began formulating policies to foster AI development. In 2018, the National Institution for Transforming India (NITI Aayog) launched the National Strategy for Artificial Intelligence, also known as #AIforAll (George, 2024). This strategy outlined India's vision to leverage AI for inclusive growth and economic development, focusing on key sectors such as healthcare, agriculture, education, smart cities, and smart mobility. The policy emphasized the need for creating a robust AI ecosystem, promoting research and innovation, and addressing ethical and privacy concerns associated with AI.

Government initiatives have played a pivotal role in promoting AI research and development. The Ministry of Electronics and Information Technology (MeitY) has been instrumental in supporting AI projects through various schemes and programs (Chatterjee, S., & NS, 2022). One such initiative is the establishment of the National AI Portal, a one-stop digital platform for AI resources, news, and

research (Chatterjee, 2020). Additionally, the government has launched the Atal Innovation Mission (AIM), which aims to create an environment of innovation and entrepreneurship, encouraging the development of AI-based solutions (Ellikkal & Rajamohan, 2023).

Several key organizations and research institutions have contributed significantly to AI advancements in India. The Indian Institutes of Technology (IITs) and Indian Institutes of Information Technology (IIITs) are at the forefront of AI research and education (George, 2024). IIT Bombay, IIT Delhi, and IIIT Hyderabad, in particular, have established dedicated AI research centers that focus on cutting-edge AI technologies and their applications (Sanyal & Chakrabarti, 2020). These institutions collaborate with industry partners and international organizations to drive AI innovation and address real-world challenges.

In addition to academic institutions, various government and private sector organizations have made substantial contributions to AI research and development. The Indian Space Research Organisation (ISRO) has integrated AI into its satellite and space exploration programs, enhancing data analysis and mission planning capabilities (Saxena, 2023). The National Association of Software and Service Companies (NASSCOM) has also played a crucial role by establishing the Centre of Excellence for Data Science and Artificial Intelligence (CoE-DSAI) (Panigrahi, 2024). This center aims to accelerate AI adoption in industry and promote collaboration between academia, industry, and government.

Furthermore, the private sector has been a key driver of AI innovation in India. Leading technology companies such as Tata Consultancy Services (TCS), Infosys, and Wipro have invested heavily in AI research and development (Chakrabarti & Sanyal, 2020). These companies collaborate with academic institutions and startups to develop AI solutions that cater to various industries, including healthcare, finance, and manufacturing.

IV. ADVANCEMENTS IN AI

Artificial Intelligence has witnessed remarkable advancements in India, transforming various sectors and driving innovation. This section explores the significant impact of AI in healthcare, agriculture, education, and other sectors, highlighting specific applications, case studies, and success stories.

AI IN HEALTHCARE

AI has revolutionized the healthcare sector in India, offering solutions that enhance diagnostic accuracy, streamline administrative processes, and improve patient outcomes. One notable application is in diagnostic imaging, where AI algorithms assist in analyzing medical images to detect diseases such as cancer, tuberculosis, and cardiovascular conditions. For example, Roy (2022) reported that AI-driven tools, like Qure.ai's qXR, have been instrumental in diagnosing tuberculosis by analyzing chest X-rays with high precision. This technology not only speeds up the diagnostic process but also ensures early detection, leading to timely treatment and better patient prognosis.

AI applications in healthcare also extend to predictive analytics and personalized medicine. Predictive analytics uses AI to forecast disease outbreaks and patient deterioration, enabling healthcare providers to take proactive measures. Personalized medicine, on the other hand, involves tailoring treatment plans based on individual patient data, such as genetic information and medical history. According to Sharma *et al*. (2022), companies like PharmEasy and 1mg are leveraging AI to offer personalized healthcare services, improving treatment efficacy and patient satisfaction.

Several case studies illustrate the impact of AI in healthcare. For instance, the AI-powered project by Wadhwani AI in Maharashtra significantly improved tuberculosis detection rates. The project used machine learning algorithms to analyze chest X-rays, enabling early diagnosis and reducing the spread of the disease. Another notable example is the use of AI by Apollo Hospitals to predict heart disease risk using patient data, which has led to better preventive care and reduced incidence of heart attacks (Agarwal *et al.*, 2022).

AI IN AGRICULTURE

The agricultural sector in India has also benefited significantly from AI, with smart farming techniques and predictive analytics playing a crucial role in enhancing productivity and sustainability. AI technologies provide farmers with real-time data on soil health, weather conditions, and crop growth, enabling them to make informed decisions and optimize resource use.

Smart farming techniques, such as precision agriculture, leverage AI to monitor and manage crops efficiently. Shaikh, Rasool, & Lone (2022) highlight the use of AI-powered drones and sensors to collect data on crop health, soil moisture, and pest infestations. This data is then analyzed to provide actionable insights, helping farmers apply the right amount of water, fertilizers, and

pesticides at the right time. As a result, farmers can increase crop yields, reduce costs, and minimize environmental impact.

Success stories in AI-driven agriculture include the Precision Farming Project by Microsoft and ICRISAT. This project uses AI models to analyze weather data, soil health, and crop conditions, providing farmers with actionable insights to optimize farming practices and improve yields (Kumar *et al.*, 2020). The project has demonstrated significant improvements in crop productivity and resource management, showcasing the potential of AI in transforming agriculture.

Another success story is the use of AI by startups like CropIn and Ninjacart. CropIn leverages AI to provide farmers with real-time data on crop health and market trends, helping them make informed decisions and maximize profits (Wolfert *et al.*, 2017). Ninjacart uses AI to streamline the supply chain, connecting farmers directly with retailers and reducing post-harvest losses (Suresh *et al.*, 2024). These initiatives have not only improved agricultural productivity but also enhanced the livelihoods of farmers.

AI IN EDUCATION

AI is transforming the education sector in India by offering personalized learning experiences, adaptive assessments, and intelligent tutoring systems. These AI-driven tools cater to the diverse needs of students, enhance engagement, and improve learning outcomes.

Personalized learning involves tailoring educational content and teaching methods to individual student needs and preferences. AI-powered platforms like BYJU's and Vedantu use machine learning algorithms to analyze student performance data and provide customized learning paths (Goswami & Sharma, 2024). These platforms offer interactive lessons, quizzes, and feedback, helping students learn at their own pace and improve their understanding of complex concepts.

AI-driven tools also support adaptive assessments, which adjust the difficulty of questions based on student performance. This approach ensures that assessments are neither too easy nor too difficult, providing an accurate measure of student knowledge and skills. According to Karan & Angadi (2023), AI-based assessment tools like Embibe are helping educators identify learning gaps and provide targeted interventions to address them.

Several examples of AI in Indian educational institutions illustrate the transformative potential of these technologies. For instance, IIT Bombay has developed an AI-powered chatbot, Asha, to assist students with academic queries and provide personalized guidance (Shrivastava & Mahajan, 2016). The chatbot

uses natural language processing (NLP) to understand student questions and deliver relevant answers, improving student support and engagement.

Another example is the use of AI by Amity University to enhance the learning experience. The university has integrated AI-powered tools into its curriculum, offering students access to intelligent tutoring systems, virtual labs, and personalized feedback. These tools have improved student performance and satisfaction, demonstrating the value of AI in education.

OTHER SECTORS

Beyond healthcare, agriculture, and education, AI has also made significant inroads in other sectors such as finance, transportation, and manufacturing.

In the finance sector, AI is being used to enhance fraud detection, automate customer service, and improve investment strategies. For example, Bajaj Finserv employs AI algorithms to detect fraudulent transactions and protect customer accounts (Patil & Kulkarni, 2019). AI-powered chatbots and virtual assistants are also being used by banks and financial institutions to provide 24/7 customer support and streamline operations.

The transportation sector has seen the adoption of AI in areas such as traffic management, autonomous vehicles, and logistics. According to Samim (2023), AI-driven traffic management systems are being implemented in cities like Bangalore and Delhi to reduce congestion and improve traffic flow. Autonomous vehicle initiatives by companies like Tata Motors and Mahindra & Mahindra are also exploring the potential of AI to enhance road safety and driving efficiency

In manufacturing, AI is being used to optimize production processes, improve quality control, and predict maintenance needs. AI-powered robots and automation systems are enhancing productivity and reducing costs in manufacturing plants. For instance, Tata Steel uses AI to monitor and analyze production data, identifying areas for improvement and ensuring high-quality output (Gajdzik & Wolniak, 2021).

V. CHALLENGES FACING AI IN INDIA

Despite the remarkable advancements in AI, several challenges impede its full potential in India. These challenges span across data privacy and security concerns, infrastructural and technological limitations, skill gaps, and regulatory and policy hurdles. Addressing these challenges is crucial for the sustainable and ethical development of AI in the country.

DATA PRIVACY AND SECURITY CONCERNS

Data privacy and security are paramount concerns in the AI landscape, given the vast amounts of personal and sensitive data that AI systems process. The adoption of AI technologies often involves collecting and analyzing data from various sources, raising significant privacy issues. Dixon (2017) emphasize that without robust data protection measures, individuals' privacy can be compromised, leading to misuse of personal information and identity theft. Moreover, the increasing instances of data breaches highlight the urgent need for enhanced security protocols.

Legal and ethical implications of data usage in AI are profound. The lack of comprehensive data protection laws in India exacerbates these concerns. The Personal Data Protection Bill, introduced in 2019, aims to address some of these issues by establishing guidelines for data collection, storage, and processing (Hoofnagle, Van Der Sloot, & Borgesius, 2019). However, its implementation has faced delays, leaving gaps in the legal framework. Ethical considerations, such as bias in AI algorithms and the potential for discrimination, also require stringent oversight to ensure AI systems are fair and unbiased.

Public perception and trust issues further complicate the adoption of AI. According to a survey by Sharma *et al.* (2022), a significant portion of the Indian population is wary of AI due to concerns about data privacy and job displacement. Building public trust in AI technologies necessitates transparent practices, robust data protection measures, and clear communication about the benefits and risks of AI.

INFRASTRUCTURAL AND TECHNOLOGICAL LIMITATIONS

Infrastructural and technological limitations present significant barriers to the widespread adoption of AI in India. One major challenge is the access to high-quality data, which is essential for training effective AI models. Data in India is often fragmented, inconsistent, and of varying quality, making it difficult to derive actionable insights. Additionally, data silos within organizations hinder the sharing and integration of data across platforms.

The need for better computational resources is another critical issue. Advanced AI algorithms require substantial computing power, which is not readily available in many parts of India. According to Lins *et al.* (2021), the lack of high-performance computing infrastructure limits the ability of researchers and businesses to develop and deploy sophisticated AI solutions. Investing in state-of-the-art data centers and cloud computing services is essential to overcome these limitations.

SKILL GAP AND WORKFORCE READINESS

The skill gap in AI and related fields is a significant hurdle in India's AI journey. While there is a growing demand for AI professionals, the supply of skilled individuals remains limited. Sharma *et al.* (2022) highlight that a considerable portion of the workforce lacks the necessary skills to develop, implement, and manage AI systems. This gap restricts innovation and slows down the adoption of AI technologies across industries.

Educational and training requirements need to be addressed to bridge this skill gap. Enhancing AI education in schools, colleges, and universities is crucial. Incorporating AI and machine learning courses into curricula, providing hands-on training, and encouraging research in AI are necessary steps. Online platforms offering AI courses and certifications, such as Coursera, edX, and Udacity, are also valuable resources for upskilling the workforce.

Several initiatives aim to bridge the skill gap in AI. The National AI Portal, launched by the Indian government, offers resources and training programs to build AI capabilities (Chatterjee, 2020). Additionally, the Atal Innovation Mission (AIM) and the NITI Aayog's AI for All initiative focus on promoting AI literacy and encouraging innovation. Collaborations between academia, industry, and government are essential to create a skilled workforce ready to harness the potential of AI.

REGULATORY AND POLICY CHALLENGES

Regulatory and policy challenges significantly impact the development and deployment of AI in India. The current regulatory framework is fragmented and lacks comprehensive guidelines specific to AI. Smuha (2021) notes that the absence of clear regulations creates uncertainty and hampers the growth of AI technologies. There is a need for policies that address ethical considerations, standardize AI practices, and promote innovation while ensuring public trust.

Current regulations, such as the Information Technology Act of 2000, provide a general framework for data protection but do not specifically address the unique challenges posed by AI (Basu & Jones, 2005). The Personal Data Protection Bill, once implemented, is expected to fill some of these gaps by establishing stricter data protection norms. However, the dynamic nature of AI technologies necessitates continuous updates to the regulatory framework to keep pace with technological advancements.

Suggestions for policy improvements include developing a comprehensive AI strategy that outlines ethical guidelines, standards for data usage, and mech-

anisms for accountability. Establishing regulatory sandboxes can allow for the testing of AI applications in a controlled environment, facilitating innovation while ensuring compliance with ethical and legal standards. Additionally, fostering international collaborations can help India align its AI policies with global best practices and stay competitive in the global AI landscape.

ADDRESSING ETHICAL CONCERNS

Ethical concerns in AI revolve around issues such as bias, fairness, and accountability. AI algorithms, if not carefully designed, can perpetuate existing biases and lead to discriminatory outcomes. Ensuring fairness in AI systems requires rigorous testing and validation to identify and mitigate biases. Transparency in AI decision-making processes is also essential to build trust and accountability.

The ethical use of AI also involves considering the societal impact of AI technologies. The potential for job displacement due to automation is a significant concern. Policies must address the need for reskilling and upskilling workers to adapt to the changing job market. Additionally, ethical frameworks should guide the development and deployment of AI in sensitive areas such as healthcare and criminal justice to prevent harm and ensure equitable outcomes.

INFRASTRUCTURE DEVELOPMENT

Developing the necessary infrastructure to support AI is critical for its widespread adoption. This includes investing in high-speed internet connectivity, particularly in rural areas, to ensure that AI technologies can be accessed and utilized by all. Enhancing data storage and processing capabilities through the establishment of modern data centers is also essential.

Public-private partnerships can play a crucial role in infrastructure development. Collaborations between the government and private sector can accelerate the deployment of AI infrastructure and ensure that resources are utilized effectively. For instance, initiatives like the Digital India program aim to improve digital infrastructure and promote digital literacy, creating a conducive environment for AI growth (Chatterjee, 2020).

PROMOTING AI RESEARCH AND INNOVATION

Promoting AI research and innovation is vital for maintaining India's competitive edge in the global AI landscape. Encouraging interdisciplinary research that combines AI with other fields, such as healthcare, agriculture, and environmental science, can lead to innovative solutions for complex problems. Providing funding and resources for AI research projects can foster a culture of innovation and drive technological advancements.

Creating innovation hubs and incubators can support startups and entrepreneurs working on AI technologies. These hubs can provide access to mentorship, funding, and networking opportunities, enabling the growth of a vibrant AI ecosystem. Additionally, fostering collaborations between academic institutions, research organizations, and industry can facilitate knowledge exchange and accelerate innovation.

VI. CASE STUDIES

AI has been applied successfully across various sectors in India, with several projects standing out for their innovative approaches and impactful results. Detailed analysis of these successful AI projects provides valuable insights and lessons that can guide future initiatives.

One notable AI project is the Wadhwani AI initiative focused on improving tuberculosis (TB) detection rates in Maharashtra. This project utilized machine learning algorithms to analyze chest X-rays, significantly enhancing the accuracy and speed of TB diagnosis. Before the implementation of AI, TB detection relied heavily on manual analysis, which was time-consuming and prone to errors (Jain, Garg, & Khera, 2022). The AI-driven approach not only accelerated the diagnostic process but also ensured early detection, which is crucial for effective treatment and control of the disease. According to the project report, the use of AI in TB detection led to a 20% increase in the identification of positive cases, highlighting the potential of AI to transform healthcare diagnostics.

Another successful AI application is seen in the agricultural sector through the Precision Farming Project by Microsoft and ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). This project leverages AI models to analyze weather patterns, soil health, and crop conditions, providing farmers with actionable insights to optimize their farming practices. For instance, AI-driven recommendations on irrigation schedules, fertilizer application, and pest management have helped farmers increase crop yields while reducing resource use. The project's impact is evident in the significant improvements

in productivity and sustainability observed in participating farms. Kumar *et al*. (2020) reported that the project led to a 30% increase in crop yields and a 20% reduction in water usage, demonstrating the transformative potential of AI in agriculture.

In the education sector, AI-powered platforms like BYJU's have revolutionized personalized learning. BYJU's uses machine learning algorithms to assess individual student performance and provide customized learning paths. The platform offers interactive lessons, quizzes, and feedback tailored to each student's learning pace and style. This personalized approach has resulted in improved engagement and learning outcomes. According to a study by Goswami & Sharma (2024), students using BYJU's showed a 25% improvement in their academic performance compared to those using traditional learning methods. This case study underscores the effectiveness of AI in enhancing educational experiences and outcomes.

AI's application in finance is exemplified by HDFC Bank's AI-driven customer service initiative. The bank implemented an AI-powered chatbot named EVA (Electronic Virtual Assistant) to handle customer queries and provide instant support. EVA uses natural language processing (NLP) to understand and respond to customer questions, offering a seamless and efficient customer service experience. Since its launch, EVA has handled millions of customer interactions, reducing response times and improving customer satisfaction. Patil & Kulkarni (2019) reported a 50% reduction in call center workload and a 20% increase in customer satisfaction scores, highlighting the benefits of AI in customer service.

LESSONS LEARNED

Several lessons can be drawn from these successful AI projects in India. Firstly, collaboration between various stakeholders, including government bodies, private sector companies, and research institutions, is crucial for the successful implementation of AI projects. The involvement of multiple stakeholders ensures access to diverse expertise, resources, and perspectives, facilitating comprehensive and effective solutions.

Secondly, the importance of context-specific AI solutions cannot be overstated. AI technologies must be tailored to address the unique challenges and needs of different sectors and regions. For example, the Precision Farming Project's success is largely attributed to its focus on the specific agricultural conditions of semi-arid regions in India.

Another key lesson is the necessity of robust data infrastructure. High-quality, reliable data is the backbone of effective AI systems. Projects like the Wadhwani AI TB detection initiative relied on extensive datasets of chest X-rays to train their machine learning models, highlighting the importance of data accessibility and quality.

Additionally, continuous monitoring and evaluation are essential to measure the impact of AI projects and identify areas for improvement. Regular assessments help in fine-tuning AI models, ensuring their accuracy and relevance over time.

Finally, addressing ethical and privacy concerns is vital for building public trust and ensuring the responsible use of AI. Transparent practices, clear communication, and adherence to data protection regulations are necessary to mitigate public concerns and promote the ethical deployment of AI technologies.

VII. FUTURE PROSPECTS AND RECOMMENDATIONS

The future prospects of AI in India are promising, with the potential to revolutionize various sectors and drive significant socio-economic growth. As AI technologies continue to advance, they are expected to become more integrated into everyday applications, offering innovative solutions to complex problems. In healthcare, future developments may include more sophisticated AI-driven diagnostic tools and personalized treatment plans that leverage genetic data and patient history. In agriculture, AI could enable precision farming on a larger scale, optimizing resource use and boosting crop yields. In education, AI-driven personalized learning platforms may become even more adaptive, offering customized educational experiences that cater to individual student needs and learning styles.

To fully realize these potential developments, strategic recommendations must be implemented to overcome current challenges. Addressing data privacy and security concerns is paramount. The implementation of robust data protection laws, such as the Personal Data Protection Bill, should be prioritized to safeguard personal information and build public trust in AI technologies. Additionally, investing in high-quality data infrastructure and computational resources is essential to support the development and deployment of advanced AI models. Enhancing internet connectivity, particularly in rural areas, will ensure that AI technologies are accessible to all, bridging the digital divide.

Another critical area is the skill gap in AI and related fields. Educational institutions should integrate AI and machine learning courses into their curricula

and provide hands-on training opportunities. Online platforms offering AI courses and certifications can complement formal education and help upskill the workforce. Collaborative efforts between the government, private sector, and academia are crucial in creating a skilled workforce ready to harness the potential of AI.

The government plays a pivotal role in fostering AI growth through policy initiatives, funding for research and development, and the creation of a conducive regulatory environment. Policies that promote innovation while addressing ethical and legal considerations will be vital in guiding the responsible development of AI. The private sector, on the other hand, can drive AI adoption by investing in research, developing AI solutions tailored to the Indian context, and fostering partnerships with academic institutions and startups.

Academic institutions are instrumental in advancing AI research and training the next generation of AI professionals. By establishing AI research centers, promoting interdisciplinary research, and collaborating with industry partners, academia can contribute significantly to AI innovation. Furthermore, creating innovation hubs and incubators can support startups and entrepreneurs working on AI technologies, providing them with the resources and mentorship needed to succeed.

VIII. CONCLUSION

This chapter has explored the significant advancements and ongoing challenges in the field of Artificial Intelligence in India, highlighting the transformative potential of AI across various sectors such as healthcare, agriculture, education, and more. The historical context provided an understanding of the early developments and government initiatives that have shaped India's AI landscape, while detailed case studies demonstrated the practical applications and impacts of AI technologies.

Despite the remarkable progress, several challenges hinder the full potential of AI in India. Data privacy and security concerns, infrastructural and technological limitations, skill gaps, and regulatory and policy hurdles present significant barriers that must be addressed. The analysis underscored the importance of robust data protection laws, investment in high-quality data infrastructure, and enhanced computational resources. Additionally, the need for a skilled workforce and comprehensive regulatory frameworks was emphasized as critical to fostering AI growth.

Future prospects for AI in India are promising, with the potential to revolutionize various sectors and drive socio-economic growth. Strategic recommendations for overcoming current challenges include prioritizing data privacy, investing in infrastructure, bridging the skill gap, and fostering a collaborative ecosystem involving the government, private sector, and academia. The role of the government in providing policy support, the private sector in driving adoption and innovation, and academia in advancing research and training is crucial in realizing the full potential of AI.

IX. BIBLIOGRAPHY

Agarwal, Y., Jain, M., Sinha, S., & Dhir, S. (2020). Delivering high-tech, AI-based health care at Apollo Hospitals. *Global Business and Organizational Excellence*, *39*(2), 20-30.

Basrur, R., & Wu, S. S. (2023). India's conventional strategy in a nuclear environment: a neglected link. *Defence Studies*, *23*(3), 457-476.

Basu, S., & Jones, R. (2005). Indian Information and Technology Act 2000: review of the regulatory powers under the Act. *International Review of Law, Computers & Technology*, *19*(2), 209-230.

Chakrabarti, R., & Sanyal, K. (2020). Towards a 'Responsible AI': Can India take the lead?. *South Asia Economic Journal*, *21*(1), 158-177.

Chakraborty, U. (2020). *Artificial Intelligence for All: Transforming Every Aspect of Our Life*. Bpb publications.

Chatterjee, S. (2020). AI strategy of India: policy framework, adoption challenges and actions for government. *Transforming Government: People, Process and Policy*, *14*(5), 757-775.

Chatterjee, S., & NS, S. (2022). Artificial intelligence and human rights: a comprehensive study from Indian legal and policy perspective. *International Journal of Law and Management*, *64*(1), 110-134.

Davenport, T. H., & Mittal, N. (2023). *All-in on AI: How smart companies win big with artificial intelligence*. Harvard Business Press.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). Technology acceptance model. *J Manag Sci*, *35*(8), 982-1003.

Dixon, P. (2017). A Failure to "Do No Harm"--India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the US. *Health and technology*, *7*(4), 539-567.

Ellikkal, A., & Rajamohan, S. (2023). Unleashing the Entrepreneurial Spirit: An Analysis of Government Programs and their Impact on Entrepreneurship Education in India. *Journal of Management*, *10*(4), 28-37.

Eubanks, B. (2022). *Artificial intelligence for HR: Use AI to support and develop a successful workforce*. Kogan Page Publishers.

Gajdzik, B., & Wolniak, R. (2021). Transitioning of steel producers to the steelworks 4.0—Literature review with case studies. *Energies*, *14*(14), 4109.

George, A. S. (2024). India's Ascent as the Global Epicenter of Artificial Intelligence. *Partners Universal Innovative Research Publication*, *2*(1), 1-15.

Goswami, A., & Sharma, A. (2024). AI For Bridging Socio-Economic Inequities in Indian Education Space. *International Journal of Research and Scientific Innovation*, *11*(4), 890-935.

Hassan, M., Malhotra, K., & Firdaus, M. (2022). Application of artificial intelligence in IoT security for crop yield prediction. *ResearchBerg Review of Science and Technology*, *2*(1), 136-157.

Hoofnagle, C. J., Van Der Sloot, B., & Borgesius, F. Z. (2019). The European Union general data protection regulation: what it is and what it means. *Information & Communications Technology Law*, *28*(1), 65-98.

Jain, R., Garg, N., & Khera, S. N. (2022). Adoption of AI-enabled tools in social development organizations in India: An extension of UTAUT model. *Frontiers in Psychology*, *13*, 893691.

Jaiswal, A., & Arun, C. J. (2021). Potential of Artificial Intelligence for transformation of the education system in India. *International Journal of Education and Development using Information and Communication Technology*, *17*(1), 142-158.

Karan, B., & Angadi, G. R. (2023). Artificial intelligence integration into school education: a review of indian and foreign perspectives. *Millennial Asia*, 09763996231158229.

Kumar, R., Yadav, S., Kumar, M., Kumar, J., & Kumar, M. (2020). Artificial intelligence: new technology to improve Indian agriculture. *International Journal of Chemical Studies*, *8*(2), 2999-3005.

Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of ambient intelligence and humanized computing*, *14*(7), 8459-8486.

Lins, S., Pandl, K. D., Teigeler, H., Thiebes, S., Bayer, C., & Sunyaev, A. (2021). Artificial intelligence as a service: classification and research directions. *Business & Information Systems Engineering*, *63*, 441-456.

Mallick, P. K. (2018). Artificial Intelligence in Armed Forces: An Analysis. *CLAWS journal*, *11*(2), 63-79.

Nedungadi, P. P., Menon, R., Gutjahr, G., Erickson, L., & Raman, R. (2018). Towards an inclusive digital literacy framework for digital India. *Education+ Training*, *60*(6), 516-528.

Panigrahi, C. M. A. (2024). Impact of artificial intelligence on Indian economy. *Panigrahi A, Ahirrao SC, Patel A, Impact of artificial intelligence on Indian economy. J Manag Res Anal*, *11*(1), 33-40.

Park, S. C. (2018). The Fourth Industrial Revolution and implications for innovative cluster policies. *Ai & Society*, *33*, 433-445.

Patil, K., & Kulkarni, M. S. (2019). Artificial intelligence in financial services: Customer chatbot advisor adoption. *Int. J. Innov. Technol. Explor. Eng*, *9*(1), 4296-4303.

Roy, R. (2022). The COVID-19 opportunity for Qure. ai. *The International Journal of Entrepreneurship and Innovation*, *23*(2), 138-143.

Samim, A. (2023). Analysis of Artificial Intelligence in Traffic Congestion and Management System. *International Journal of Science and Research (IJSR)*, *12*(1), 679-684.

Sanyal, K., & Chakrabarti, R. (2020). *Artificial Intelligence and India*. Oxford University Press.

Saxena, A. (2023). India's Space Policy and Counter-Space Capabilities. *Strategic Analysis*, *47*(2), 146-158.

Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, *198*, 107119.

Sharma, A., Sharma, R. K., Nanda, S., & Misra, A. (2022). Emerging real time analytics based health start-ups: opportunities during covid-19. *Journal of Medical Pharmaceutical and Allied Sciences*, *11*(4), 5017-5025.

Sharma, M., Luthra, S., Joshi, S., & Kumar, A. (2022). Implementing challenges of artificial intelligence: Evidence from public manufacturing sector of an emerging economy. *Government Information Quarterly*, *39*(4), 101624.

Shrivastava, R., & Mahajan, P. (2016). Artificial intelligence research in India: a scientometric analysis. *Science & Technology Libraries*, *35*(2), 136-151.

Smuha, N. A. (2021). From a 'race to AI'to a 'race to AI regulation': regulatory competition for artificial intelligence. *Law, Innovation and Technology*, *13*(1), 57-84.

Solove, D. J. (2005). A taxonomy of privacy. Univ. Pa. Law Rev. 154(3), 477–560

Suresh, D., Choudhury, A., Zhang, Y., Zhao, Z., & Shaw, R. (2024). The Role of Data-Driven Agritech Startups—The Case of India and Japan. *Sustainability*, *16*(11), 4504.

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming–a review. *Agricultural systems*, *153*, 69-80.

CAPÍTULO SIETE

INTELIGENCIA ARTIFICIAL. LOS NUEVOS POTENCIALES EN SUPERVISIÓN Y SEGURIDAD PÚBLICA

INTELIGENCIA ARTIFICIAL, DESINFORMACIÓN Y SEGURIDAD PÚBLICA EN AMÉRICA LATINA

SUMARIO

I. Introducción **II.** La inteligencia artificial en tiempos de la desinformación **III.** Inteligencia Artificial y Seguridad Pública en América Latina **IV.** Conclusión y hallazgos **V.** Bibliografía

> FAUSTO CARBAJAL GLASS

Senior Policy Advisor en Miranda Partners. Socio fundador de Delphi Solutions and Strategic Services. Consultor en riesgo político y seguridad. Internacionalista por la Universidad Iberoamericana y Maestro en Estudios de Guerra por el King's College London.

> AMAURY FIERRO GONZÁLEZ

Especialista en inteligencia y geopolítica. Actualmente es instructor en Rabdan Academy, la más importante institución educativa en temas de defensa y seguridad en Medio Oriente, y director de aprendizaje en Tech Corp.

I. INTRODUCCIÓN

La irrupción de la inteligencia artificial (IA) ha transformado profundamente la manera en que las sociedades se organizan, comunican e interactúan. En América Latina y el Caribe, esta transformación presenta tanto oportunidades como desafíos únicos, particularmente en dos ámbitos críticos para la gobernabilidad democrática y el desarrollo social, a saber: la desinformación y la seguridad pública. Estos fenómenos, aunque distintos en su naturaleza, comparten una característica fundamental, es decir, a su capacidad para erosionar la confianza ciudadana en las instituciones y alterar el tejido social de las comunidades.

La región latinoamericana enfrenta una coyuntura particular en la que convergen múltiples crisis: institucional, de seguridad, de legitimidad política y de cohesión social. En este contexto, la IA emerge como una herramienta de doble filo. Por un lado, ofrece soluciones innovadoras para combatir problemas arraigados. Por otra parte, su implementación inadecuada puede exacerbar las desigualdades existentes y crear nuevas vulnerabilidades a sociedades enteras. Esta dualidad se manifiesta claramente en los dos ejes temáticos que aborda este capítulo.

En el ámbito de la desinformación, América Latina se ha convertido en un terreno particularmente fértil para la proliferación de noticias falsas y contenido manipulado. Como se verá más adelante, la penetración masiva de redes sociales, combinada con bajos niveles de alfabetización digital y contextos de alta polarización política, ha creado condiciones ideales para la difusión de desinformación. La IA, a través de tecnologías como los *deepfakes* y la generación automatizada de contenido, ha amplificado estos desafíos. Sin embargo, también proporciona herramientas sofisticadas para la detección y

el combate de la desinformación, desde algoritmos de verificación automática hasta sistemas de análisis de patrones de difusión. Así, como mejor juzgará el lector, el presente capítulo también querrá enfocarse en el lado amable de la IA.

En materia de seguridad pública, la región enfrenta retos complejos que incluyen altas tasas de criminalidad, crimen organizado transnacional y violencia homicida endémica en ciertas zonas. La IA ofrece soluciones prometedoras a través de sistemas predictivos, videovigilancia inteligente y análisis de patrones delictivos. No obstante, la implementación de estas tecnologías plantea dilemas significativos en torno a la privacidad, los derechos civiles y el uso ético de la vigilancia masiva. La tentación del "solucionismo tecnológico" –la creencia de que la tecnología por sí sola puede resolver problemas sociales complejos– es particularmente fuerte en este ámbito, y el capítulo también retomará esta preocupación.

De manera que el capítulo examina la intersección de la IA con estos dos fenómenos críticos para América Latina, analizando tanto las innovaciones prometedoras como los riesgos inherentes a su implementación. La primera sección aborda cómo la IA está transformando el panorama de la desinformación, explorando tanto su papel en la creación de contenido engañoso como su potencial para combatirlo. La segunda sección analiza la aplicación de la IA en seguridad pública, examinando casos de implementación en diversos países de la región y evaluando su impacto en la efectividad policial y los derechos ciudadanos.

El análisis se fundamenta en casos concretos de implementación en diversos países latinoamericanos, evidenciando cómo diferentes contextos sociopolíticos y marcos regulatorios conducen a resultados distintos. Asimismo, se consideran las implicaciones *éticas y sociales de estas tecnologías, reconociendo que su implementación no ocurre en un vacío, sino en sociedades marcadas por profundas desigualdades y desafíos estructurales.*

El objetivo último es proporcionar una comprensión balanceada de cómo la IA puede contribuir a abordar dos de los desafíos más apremiantes que enfrenta América Latina, sin perder de vista la necesidad de un enfoque integral que considere las dimensiones sociales, éticas y políticas de su implementación. Solo a través de este entendimiento holístico será posible aprovechar el potencial transformador de la IA mientras se mitigan sus riesgos y se protegen los derechos fundamentales de los ciudadanos.

II. LA INTELIGENCIA ARTIFICIAL EN TIEMPOS DE LA DESINFORMACIÓN

La relación del ser humano con la verdad ha sido especialmente compleja a lo largo de la historia. Acaso un poco más hoy en día con las redes sociales y la inteligencia artificial (IA). No porque necesariamente la hayamos alcanzado antes, sino porque pareciera que nos estamos alejando de ella en la actualidad. Nuestra relación con la verdad actualmente no podría entenderse sin considerar las ambigüedades del mundo VICA –más volátil, incierto, complejo y ambiguo. Todo puede ser o no ser, en función de nuestras creencias, paradigmas y, sobre todo, emociones.[1] Más aún, en tiempos de desinformación y *fake news*, podría decirse que ya no se trata de la búsqueda de la verdad en sí mismo, sino el impulso de reafirmar nuestros propios sesgos cognitivos.[2] Y a un paso de distancia se encuentra el relativismo.

Lo preocupante de esta forma de relacionarnos con la información –y con la verdad– es que tiene implicaciones directas y tangibles en las sociedades, particularmente las democráticas. Por ejemplo, el uso político de la inteligencia artificial (IA), a través de la creación y difusión de *deepfakes* y otros contenidos fabricados tendenciosamente, hace cada vez más difícil para los ciudadanos discernir la verdad de lo falso, lo que inevitablemente debilita instituciones y socava el proceso democrático.[3] Tal como lo señalara Timothy Snyder: el problema empieza cuando la verdad y los datos duros se vuelven irrelevantes en una sociedad que ha dejado de ser fáctica y, en cambio, se ha convertido en una sociedad oracular.[4] Esta transición de sociedades fácticas a oraculares, ha

1 Para un estudio sobre el componente emocional de la posverdad en las redes sociales, ver: Durnová, Anna. 2019. "Unpacking emotional contexts of post-truth". *Critical Policy Studies*, 13 (4), 447-450.

2 De acuerdo con la literatura en la materia, la redes sociales exacerban primordialmente cinco sesgos cognitivos, a saber: rebaño, encuadre, exceso de confianza, confirmación y anclaje. Ver: French, Aaron M.; Storey, Veda C. & Wallace, Linda. 2023. "The impact of cognitive biases on the believability of fake news". *European Journal of Information Systems*, 1 – 22.

3 Labuz, Mateusz & Nehring, Christopher. 2024. "On the way to deep fake democracy? Deep fakes in election campaigns in 2023". *European Political Science*, DOI https://doi.org/10.1057/s41304-024-00482-9; Diakopoulos, Nicholas & Johnson, Deborah. 2020. "Anticipating and addressing the ethical implications of deepfakes in the context of elections". *New Media & Society*, 23 (7), 2072 – 2098. https://journals.sagepub.com/doi/abs/10.1177/1461444820925811.

4 Snyder, Timothy. 2017. *On Tiranny: Twenty Lessons from the Twentieth Century*. Nueva York: Crown Publishing.

sentado las bases para que la desinformación –es decir, la difusión deliberada y sistemática de mentiras[5]– esté presente en los principales eventos nacionales e internacionales.

De hecho, el Reporte de Riesgos Globales de 2024, del Foro Económico Mundial (WEF, por sus siglas en inglés), identifica a la desinformación como uno de los principales riesgos que enfrenta el mundo precisamente por su severidad en el corto y largo plazos.[6] La preocupación que genera la desinformación se justifica con múltiples casos en la historia reciente. Ya sea durante el período más intenso de la Covid-19, la invasión de Rusia a Ucrania, los flujos migratorios hacia Estados Unidos, el conflicto entre la organización terrorista Hamás e Israel, la competencia estratégica entre Estados Unidos – China, o la creciente inestabilidad política y conflictividad social en América Latina y el Caribe.

Cabe señalar que, independientemente del caso, la desinformación se inserta en escenarios un patrón similar: una situación atípica o de crisis, en un entorno caracterizado por una fuerte competencia por el poder, en el que múltiples actores están interesados en ganar las mentes y los corazones de la población-audiencia. La desinformación, además, ha ocasionado que cada vez sea más difícil conocer sobre una situación específica y, más importante aún, anticipar sus resultados o implicaciones.[7]

Al respecto, es menester resaltar que la desinformación, potencializada en el mundo digital mediante la IA, forma parte esencial de lo que George Kennan denominó "*organized political warfare*", a mediados del siglo XX.[8] Más recientemente, el binomio desinformación-inteligencia artificial se ha utilizado como parte del abanico de tácticas en expresiones más nuevas y sutiles de conflicto o estados intermedios de paz y guerra, comúnmente llamadas guerras híbridas o confrontaciones de zona gris.[9] En términos generales, éstas

5 Broda, Elena & Strömbäck, Jesper. 2024. "Misinformation, disinformation, and fake news: lessons from an interdisciplinary, systematic literature review". *Annals of the International Communication Association*, 48 (2), 139 – 166.

6 WEF. 2024. "Global Risks Report 2024". Disponible en: https://www.weforum.org/publications/global-risks-report-2024/, p. 8.

7 Li, Cathy & Callegari, Agustina. 2024. "Stopping AI disinformation: Protecting truth in the digital world". *World Economic Forum*. Disponible en: https://www.weforum.org/agenda/2024/06/ai-combat-online-misinformation-disinformation/.

8 Kennan, George F. 1948. "The inauguration of organized political warfare". *Wilson Center*. Disponible en: https://digitalarchive.wilsoncenter.org/document/george-f-kennan-inauguration-organized-political-warfare.

9 Arreola García, Adolfo. 2024. "Inteligencia Artificial y Desinformación: Papel de los Conflictos del Siglo XXI". *Centro de Estudios Estratégicos del Ejército de Perú*, 3 (3), 87-113.

nuevas confrontaciones abarcan esfuerzos sistemáticos por socavar la estabilidad, integridad y permanencia de un Estado, mediante operaciones de información y *fake news*, diplomacia coercitiva, recursos legales (*lawfare*), infiltración económica, subversión política, chantajes comerciales, ataques cibernéticos, además de las acciones militares tradicionales, tanto convencionales como no convencionales –lo que incluye la movilización de actores no estatales violentos, tales como grupos criminales, subversivos o terroristas.[10]

POTENCIAL DE IA PARA COMBATIR LA DESINFORMACIÓN

Por ello, es de la mayor importancia para las sociedades democráticas revisar y componer su relación con la verdad, especialmente en medio de transiciones geopolíticas globales, recesiones democráticas a nivel global y sociedades crecientemente polarizadas. En este sentido, si bien es cierto que los avances tecnológicos, en particular la IA, han amplificado la desinformación hoy en día, a continuación se resalta el potencial que la IA tiene para combatirla.

Como se ha señalado, la proliferación de la IA en la era digital ha marcado el comienzo de innovaciones notables y en igual medida desafíos únicos, particularmente en el ámbito de la integridad de la información.[11] Sobre todo si consideramos que la inteligencia artificial generativa (GAI, por sus siglas en inglés) le añade una nueva dimensión al problema de la desinformación.[12] Aun así, es igualmente cierto que la IA juega un rol preponderante al momento de contrarrestar la desinformación.[13] Por ejemplo, los sistemas avanzados de *machine learning* impulsados por IA pueden analizar los patrones, el uso del lenguaje y contexto para ayudar en la moderación del contenido, la verificación de hechos y la detección de información falsa.[14] En este sentido, los algoritmos de la IA

10 Ball, Joshua. 2022. "The changing face of conflict: What is hybrid warfare?". *Global Security Review*. Disponible en: https://globalsecurityreview.com/hybrid-and-non-linear-warfare-systematically-erases-the-divide-between-war-peace/

11 Li & Callegari, 2024.

12 Endert, Julius. 2024. "Generative AI is the ultimate disinformation amplifier". *Deutsche Welle*. Disponible en: https://akademie.dw.com/en/generative-ai-is-the-ultimate-disinformation-amplifier/a-68593890

13 Moreno Espinosa, Pastora; Alsarayreh, Abdulsalam & Figuereo Benítez, Juan Carlos. 2024. "El *big data* y la inteligencia artificial como soluciones a la desinformación". *Doxa Comunicación*, 38.

14 Li & Callegari, 2024.

permiten identificar patrones y detectar noticias falsas, lo que ayuda al momento de la verificación de hechos (o *fact-checking*) de manera más eficiente.[15]

Al respecto, una de las formas más tangibles en que la IA puede combatir la desinformación es a través de sus sistemas de análisis de la información. De acuerdo con un estudio de la Escuela de Administración Rady de la Universidad de California San Diego, los algoritmos de aprendizaje automático (*machine learning*) superan significativamente el juicio humano al momento de detectar mentiras o *fake news* durante interacciones estratégicas de alto riesgo.[16] No se omite señalar que estudios como éstos abren la puerta para que la IA pueda realizar contribuciones significativas al momento de combatir la desinformación, dado que el *machine learning* podría ser usado para reducir el número de contenido falso en las principales redes sociales, a saber, Facebook, You Tube, Instagram y Tik-Tok.

Adicionalmente, aunque los bots se han hecho conocidos generalmente por su capacidad para desinformar, estos instrumentos también pueden contribuir a contrarrestar la desinformación. Concretamente, se trata de que "los *bots* que, impulsados por la IA, desempeñan un papel clave al proporcionar respuestas automáticas a consultas comunes que permiten difundir informaciones precisas".[17] De hecho, ya algunos gobiernos utilizaron estos *chatbots* para combatir la desinformación durante la pandemia de COVID-16. Cabe señalar que los países de América Latina y el Caribe bien podrían utilizar este tipo de recursos para mitigar los efectos de la desinformación en una sociedad particularmente afectada por este fenómeno.

En este sentido, el uso de la IA para combatir la desinformación en la región es esencial para revertir la erosión democrática por la que atraviesan diversos países del Hemisferio Occidental. Al respecto, América Latina y el Caribe es un claro ejemplo del alcance y la profundidad de la desinformación. Un alto porcentaje de ciudadanos de países latinoamericanos percibe que se enfrenta

15 Soprano, Michael; Roitero, Kevin; La Barbera, David; Ceolin, David; Spina, Damiano; Demartini, Gianluca & Mizzaro, Stefano. 2024. "Cognitive Biases in Fact-Checking and Their Countermeasures: A Review". *Information Processing and Managment*, 61, 1 – 29.

16 Serra-García, Marta & Gneezy, Uri. 2023. "Improving human deception detection using algorithmic feedback". *CESIFO Working Papers*, 10518 (2023). Disponible en: https://www.cesifo.org/en/publications/2023/working-paper/improving-human-deception-detection-using-algorithmic-feedback

17 Alonso-González, Marián & Sánchez Gonzales, Hada M. 2024. "Inteligencia artificial en la verificación de la información política. Herramientas y tipología". *Revista Más Poder Local*, 56, p. 30.

a noticias engañosas o falsas todos o casi todos los días. Brasil encabeza la lista con el 60%, seguido por Ecuador con el 58%. Adicionalmente, ciudadanos de Colombia, Chile y Argentina también experimentan altos niveles de exposición a la desinformación, con 53%, 52% y 51%, respectivamente.[18] De acuerdo con el mismo estudio, México acumuló el 43% de sus ciudadanos.

INICIATIVAS PÚBLICO-PRIVADAS

Como se ha señalado, América Latina y el Caribe es una de las regiones donde más se consumen noticias falsas y desinformación.[19] Lo anterior hace impostergable que la región adopte medidas concretas para combatir este flagelo que, como se ha dicho, atenta en contra de la gobernabilidad democrática de los países. En este sentido, conviene señalar como desde la educación algunas empresas como Google, a través de su Google News Initiative, realizan acciones orientadas a desarrollar las habilidades digitales de las personas, especialmente con el apoyo de la creación de la Red Latinoamericana de Formadores en Fact checking, en 2022. En particular, se trata de un proyecto cuyo objetivo es la formación en enseñanza de fact checking para que periodistas y profesores universitarios de la Argentina, México, Colombia y Perú actúen como multiplicadores.[20]

Por otra parte, diversas iniciativas se han enfocado específicamente en abordar el problema de la desinformación utilizando inteligencia artificial. Un ejemplo destacado es la Coalición para la Procedencia y Autenticidad del Contenido (C2PA), conformada por empresas tecnológicas líderes como Adobe, Arm, Intel, Microsoft y TruePic. Su objetivo principal es combatir la proliferación de información engañosa en línea mediante el desarrollo de estándares técnicos que permitan certificar el origen, historial y procedencia de los contenidos multimedia.[21] Otra iniciativa global es la *AI Governance Alliance*, una iniciativa

18 ACTIVA & Worlwide Independent Network of Market Research. 2022. "*Fake news: Desinformación en Chile y LatAm*". Disponible en: https://chile.activasite.com/wp-content/uploads/2023/04/222287_Estudio-Desinformacio%CC%81n_0329_V1-2.pdf

19 Duchiade, André. 2023. "Nuevo mapa destaca leyes contra la desinformación en América Latina y el Caribe". *LatAm Journalism Review*. Disponible en: https://latamjournalismreview.org/es/articles/nuevo-mapa-destaca-leyes-contra-la-desinformacion-en-america-latina-y-mas-alla/

20 Rosenzvit, Milena. 2022. "Lanzamos la Red latinoamericana de formadores en fact checking". *Chequeado*. Disponible en: https://chequeado.com/lanzamos-la-red-latinoamericana-de-formadores-en-fact-checking/

21 Para más información sobre la iniciativa, ver: Coalition for Content Provenance and Authenticity. Disponible en: https://c2pa.org/

emblemática del Foro Económico Mundial y parte del Centro para la Cuarta Revolución Industrial, misma que convoca a expertos y organizaciones alrededor del mundo para abordar los desafíos de la IA, incluida la generación de contenido engañoso o dañino y la violación de los derechos de propiedad intelectual.[22]

DESINFORMACIÓN Y POLÍTICA

Mención aparte merece el impacto que la desinformación tiene en la política e instituciones, y cómo la IA puede contribuir a la reversión de esta problemática. Más aún, estos indicadores reflejan otra problemática asociada a las noticias falsas y otras expresiones de desinformación: el desinterés y la desconfianza con que los ciudadanos se relacionan con las noticias periodísticas de medios tradicionales.[23] Desde luego, un flujo informativo deficiente ocasionará eventualmente que se incremente la desconfianza de la ciudadanía en las instituciones política y los procesos electorales –componentes esenciales en una democracia duradera.[24]

De ahí que la IA ayuda a detectar y combatir la desinformación en la política. Los algoritmos avanzados y las técnicas de aprendizaje automático permiten evaluar la credibilidad de las fuentes y prever la desinformación.[25]

Esta nueva realidad ha dado origen incluso a una nueva forma de comunicación política algorítmica.[26] Es decir, la intersección entre la computación científica y las ciencias sociales, y que se caracteriza por la microsegmentación, por los diagnósticos personalizados, por la generación automatizada de contenidos.[27] El impacto de esto es que se fomenta la creación de cámaras de eco en las que

22 Li & Callegari, 2024.

23 Alonso-González, Marián & Sánchez Gonzales, Hada M. 2024. "Inteligencia artificial en la verificación de la información política. Herramientas y tipología". *Revista Más Poder Local*, 56, p. 30.

24 Bennet, Lance & Livingston, Steven. 2018. "The disinformation order: Disruptive communication and the decline of democratic institutions". *European Journal of Communication*, 32 (2), 122 – 139.

25 Sánchez Gonzales, Hada. 2022. "Transformación digital y audiencia. Tendencia y uso digital en medios verificadores". Ámbitos. Revista Internacional de Comunicación, (56), 9 – 20.

26 Campos-Domínguez, Eva & García-Orosa, Berta. 2018. "Comunicación algorítmica en los partidos políticos: automatización de producción y circulación de mensajes". *Profesional de la Información*, 27 (4), 769 – 777.

27 López-López, Paulo Carlos; Lagares Díez, Nieves & Puentes-Rivera, Iván. 2021. "La inteligencia artificial contra la desinformación: una visión desde la comunicación política". *Razón y Palabra*, 24 (111), 5 – 11.

las personas reafirman sus ideas preconcebidas y se alteran los hechos. Para ello, se han creado múltiples herramientas de la IA que permiten contrarrestar la desinformación en las redes sociales.[28]

La comunicación política algorítmica es una problemática muy presente en América Latina y el Caribe. De hecho, Bradshaw y Howard han encontrado evidencias de un uso automatizado y humano de estrategias de manipulación en diversos países de la región principalmente en Colombia, Ecuador y México.[29] De hecho, en México se han encontrado *ciborgs*, los cuales consisten en bots o sistemas automatizados que pueden simular el comportamiento humano.[30]

Sin embargo, la IA aplicada al periodismo se posiciona como un remedio potencial para contrarrestar la desinformación, por lo menos en tres ejes: 1) la mejora del estado del periodismo profesional, 2) el manejo de la privacidad y 3) el planteamiento ético del desarrollo tecnológico, que contribuya a que las decisiones informativas sean conscientes y alejadas de los sesgos.[31] Así, la IA tiene el potencial de contribuir al desarrollo de lo que Adela Cortina llamó la "ciudadanía mediática".[32]

En este sentido, algunos académicos han sugerido un enfoque híbrido en el que las herramientas de Inteligencia Artificial y los humanos –incluidos expertos y no expertos reclutados en plataformas de crowdsourcing– trabajen juntos para combatir la difusión desinformación en internet.[33] En este sentido, de acuerdo con estos estudios, la combinación de estos métodos no sólo puede dar como resultado procesos de verificación de datos más eficientes y efectivos, sino que

28 Atanasova, Pepa; Simonsen, Jakob Grue; Lioma, Christina & Augenstein, Isabelle. 2020. "Generating fact checking explanations". In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7352–7364.

29 Bradshaw, Samantha & Howard, Philip N. 2017. *Troops, trolls and troublemakers:a global inventory of organized social media manipulation*. Oxford: University of Oxford.

30 Barredo-Ibáñez, Daniel; De la Garza Montemayor, Daniel Javier; Torres Toukoumidis, Ángel; López-López, Paulo Carlos 2021. "Artificial intelligence, communication, and democracy in Latin America: a review of the cases of Colombia, Ecuador, and Mexico". Profesional de la Información, 30 (6), 1 – 17.

31 Manfredi Sánchez, Juan Luis & Ufarte Ruiz, María José. 2020. "Inteligencia artificial y periodismo: una herramienta contra la desinformación". *Revista CIDOB d'Afers Internacionals*, 124, 49 – 72.

32 Cortina, Adela. 2004. Ciudadanía mediática. *El País*. Disponible. https://elpais.com/diario/2004/11/24/opinion/1101250805_850215.html.

33 Demartini, Gianluca; Mizzaro, Stefano & Spina, Damiano. 2020. "Human-in-the-loop Artificial Intelligence for Fighting Online Misinformation: Challenges and Opportunities". *IEEE*, 65-75.

también puede conducir a una mayor confianza en los resultados en comparación con los métodos puramente basados en IA y también puede aprovechar la dimensión humana integrada para aumentar el nivel de transparencia en torno a la clasificación de noticias falsas.

La relación contemporánea entre la inteligencia artificial, la verdad y la desinformación representa un desafío complejo para las sociedades democráticas, especialmente en América Latina y el Caribe. Mientras la IA ha amplificado los riesgos de desinformación, simultáneamente ofrece herramientas prometedoras para combatirla. La región latinoamericana se sitúa como un laboratorio para comprender los impactos de la desinformación, nos sólo para saber su profundidad y alcance, sino también posibles medidas para contrarrestarla efectivamente.

El potencial de la IA para contrarrestar este fenómeno es multidimensional. Los sistemas de *machine learning* pueden analizar patrones lingüísticos y contextuales con mayor eficiencia que el juicio humano para detectar noticias falsas. Además, emergen iniciativas público-privadas por demás prometedoras. Paralelamente, espacios como el Foro Económico Mundial convoca expertos globales para abordar los desafíos éticos de la inteligencia artificial. Un enfoque híbrido se perfila como una estrategia efectiva: combinar herramientas de IA con intervención humana, aprovechando la eficiencia tecnológica y la capacidad de contextualización de los expertos. Este modelo podría aumentar la transparencia en la clasificación de noticias falsas y generar procesos de verificación más robustos.

Por otra parte, la dimensión política de la desinformación es crucial. Es innegable que la desinformación no solo distorsiona la percepción pública, sino que socava instituciones democráticas, erosiona la confianza ciudadana y facilita la manipulación en contextos de alta polarización y competencia por el poder. Por lo tanto, la solución no radica únicamente en tecnología, sino en un enfoque integral que incluya educación digital, desarrollo de habilidades críticas de verificación y un marco ético para el uso de inteligencia artificial.

En síntesis, la IA se presenta simultáneamente como problema y solución en la lucha contra la desinformación, requiriendo una aproximación sofisticada, interdisciplinaria y colaborativa para proteger la integridad de los cimientos democráticos de las sociedades en América Latina y el Caribe, en donde este flagelo está particularmente presente.

III. INTELIGENCIA ARTIFICIAL Y SEGURIDAD PÚBLICA EN AMÉRICA LATINA

La intersección entre la inteligencia artificial (IA) y la seguridad pública representa uno de los campos más controvertidos en la implementación de tecnologías emergentes, sobre todo por la divergencia que se da entre las motivaciones detrás de su uso y la funcionalidad y usabilidad que puede aportar al ejercicio de la seguridad pública.

Por un lado, está la aspiración de los gobiernos por incorporar dispositivos tecnológicos para la monitorización de sus gobernados, que si bien no es nueva, ha encontrado en las capacidades de la IA un territorio fértil para su expansión. Por el otro, las amenazas a la seguridad pública son cada vez más dinámicas y complejas, y la integración de nuevas tecnologías puede contribuir de manera significativa al mejor combate y persecución de dichas amenazas. En esta sección se discuten varios aspectos sobre cómo se ha dado la relación entre ambas variables, primero de manera general y luego en el contexto latinoamericano.

PARADIGMAS Y APLICACIONES DE LA IA EN SEGURIDAD PÚBLICA

La integración de la IA en el ámbito de la seguridad pública se manifiesta diversos paradigmas que están influyendo, o inclusive modificando, algunas de las formas en que se concibe y ejecuta la labor policial. En la primera Reunión Global sobre las Oportunidades y Riesgos de la Inteligencia Artificial y Robótica para la Aplicación de la Ley, organizada por el Centro de Innovación de INTERPOL y el Instituto Interregional de las Naciones Unidas para Investigaciones sobre la Delincuencia y la Justicia (UNICRI) en 2018, se plantearon algunos de éstos paradigmas, y podemos identificar al menos cuatro de éstos.[34]

34 "Artificial Intelligence and Robotics for Law Enforcement." Interpol y United Nations Interregional Crime and Justice Research Institute (UNICRI), 2019. Disponible en: https://unicri.it/sites/default/files/2019-11/Artificial%20Intelligence%20and%20Robotics%20for%20Law%20Enforcement.pdf.

El primero corresponde al *machine learning* y *deep learning*[35] para el análisis predictivo del crimen. Este paradigma se centra en la identificación de patrones delictivos mediante el procesamiento de grandes volúmenes de datos históricos, permitiendo anticipar posibles incidentes y optimizar la distribución de recursos policiales.

El segundo paradigma abarca los sistemas de videovigilancia inteligente, que incluyen capacidades de reconocimiento facial, de voz y de objetos. Estos sistemas funcionan como "segundos ojos" en tareas que demandan alta precisión en la observación, siendo capaces de identificar rostros incluso en condiciones subóptimas o cuando están parcialmente cubiertos. La pandemia de COVID-19 impulsó significativamente el desarrollo de estas tecnologías, principalmente en India y China, y muchas de estas innovaciones están siendo adoptadas por corporaciones policiales latinoamericanas.

El tercer paradigma comprende las tecnologías de exploración, desde vehículos autónomos hasta análisis geoespacial. Estas herramientas permiten cubrir territorios extensos o de difícil acceso, proporcionando datos cruciales para la toma de decisiones en tiempo real. Los sistemas de análisis geoespacial integran múltiples capas de información, desde patrones de movimiento poblacional hasta incidencia delictiva, permitiendo identificar correlaciones que podrían pasar desapercibidas mediante métodos tradicionales. En el caso de los vehículos autónomos, principalmente drones y robots de exploración, estos pueden acceder a zonas de alto riesgo o difícil acceso, recopilando información visual y sensorial que aumenta la seguridad del personal policial y mejora la calidad de la información disponible para la toma de decisiones operativas.

El cuarto paradigma se refiere a las integraciones con infraestructura y arquitectura de sistemas de comunicación, fundamental para la coordinación efectiva

35 *Machine learning* (aprendizaje – máquina) y *deep learning* (aprendizaje profundo) son dos de los tipos de inteligencia que se clasifican dentro de la IA estrecha o débil (Narrow AI), el único tipo de IA que existe actualmente. La IA estrecha se especializa en tareas específicas y, aunque puede superar al ser humano en su ámbito definido, no puede operar fuera de él. Otros tipos de IA, aún teóricos, incluyen la Inteligencia Artificial General (AGI o IA fuerte), que podría igualar la capacidad cognitiva humana en cualquier tarea, y la Superinteligencia Artificial (ASI), que superaría las capacidades cognitivas humanas en todos los aspectos. En términos de funcionalidad, la IA actual se puede categorizar en IA reactiva, que opera solo con datos presentes sin memoria del pasado (como Deep Blue, el sistema de ajedrez de IBM), e IA de memoria limitada, que puede utilizar datos históricos para mejorar sus decisiones (como los sistemas de conducción autónoma). Existen también categorías teóricas como la IA con Teoría de la Mente, que comprendería emociones y pensamientos humanos, y la IA autoconsciente, que desarrollaría su propia forma de consciencia.

entre diferentes unidades y jurisdicciones. Estos sistemas facilitan no solo la transmisión de información en tiempo real entre diferentes cuerpos policiales y centros de comando, sino que también permiten la integración de múltiples fuentes de datos, desde llamadas de emergencia hasta *feeds* de videovigilancia. La arquitectura de estos sistemas incorpora capacidades de procesamiento distribuido y análisis en tiempo real, permitiendo que la información crítica llegue de manera oportuna a los equipos de respuesta. Además, estos sistemas pueden adaptarse a situaciones de crisis donde la infraestructura convencional de comunicaciones podría verse comprometida, garantizando la continuidad operativa en situaciones críticas.

Las aplicaciones específicas de estos paradigmas son vastas y diversas. En el ámbito forense, estas tecnologías han revolucionado la investigación criminal mediante herramientas avanzadas de autopsia virtual que ayudan a determinar causas de muerte,[36] sistemas de recreación de escenas del crimen,[37] y análisis automatizado de ADN.[38] La gestión del tráfico urbano se ha transformado con sistemas capaces de identificar faltas de tránsito, gestionar multas de manera automatizada y detectar vehículos robados mediante software de visión computacional.[39]

En el campo de la investigación criminal, la IA ha demostrado particular utilidad en la detección de patrones anómalos en transacciones financieras, facilitando la identificación de actividades ilícitas.[40] Los sistemas de reconocimiento de voz y análisis de audio permiten la detección y localización de disparos, proporcionando respuestas más rápidas a incidentes violentos.[41] La tecnología de lectura e identificación de placas vehiculares, una de las primeras aplicaciones

36 O'Sullivan, Shane, Andreas Holzinger, Kurt Zatloukal, Paulo Saldiva, Mohammed Imran Sajid, and Dominic Wichmann. "Machine learning enhanced virtual autopsy." Autopsy & case reports 7, no. 4 (2017): 3.

37 Sacco, M. A., P. Tarzia, L. Tarda, R. La Russa, F. Cordasco, and I. Aquila. "The artificial intelligence in autopsy and crime scene analysis." *La Clinica Terapeutica* 175, no. 4 (2024).

38 Alketbi, Salem K. "Emerging Technologies in Forensic DNA Analysis." *Available at SSRN 4956325* (2024).

39 Šusteková, D., and Marta Knutelska. "How is the artificial intelligence used in applications for traffic management." *Machines. Technologies. Materials*. 9, no. 10 (2015): 49-52.

40 Stears, Chris, and Joshua Deeks. "The use of artificial intelligence in fighting financial crime, for better or worse?." *Journal of Money Laundering Control* 26, no. 3 (2023): 433-435.

41 https://www.soundthinking.com/spanish-lp/

de redes neuronales en seguridad, ha evolucionado hasta alcanzar niveles de precisión previamente inimaginables.[42]

Una aplicación particularmente relevante es el desarrollo de herramientas para la protección de menores y la detección de fraude en línea. Por ejemplo, sistemas como PhotoDNA, desarrollado por Microsoft y utilizado por agencias de seguridad y plataformas digitales a nivel global, emplean *hashing* criptográfico y aprendizaje profundo para identificar, rastrear y eliminar imágenes de explotación infantil incluso cuando han sido alteradas. Estas herramientas pueden procesar millones de imágenes por segundo y tienen una tasa de precisión superior al 99.5% en la identificación de material ilegal.[43] En el ámbito del fraude financiero en sus múltiples modalidades, un ejemplo es *Feedzai*, una compañía portuguesa que utiliza IA para analizar más de 100 mil millones de transacciones anuales, detectando patrones de comportamiento sospechoso en tiempo real y reduciendo las pérdidas por fraude hasta en un 75%. Esta compañía se encuentra colaborando con bancos brasileños y mexicanos, así como con sus respectivas autoridades regulatorias, para poder combatir este ilícito.[44]

Los sistemas de trazado de paquetes basados en criptografía representan un avance significativo en la investigación digital, permitiendo a las fuerzas del orden rastrear actividades ilícitas mientras preservan la privacidad de los usuarios legítimos. Por ejemplo, tecnologías como la encriptación homomórfica permiten analizar datos cifrados sin necesidad de descifrarlos primero. Sistemas como *CryptoGraph* utilizan esta técnica para monitorear patrones de tráfico en la red y detectar actividades maliciosas, mientras mantienen el contenido de las comunicaciones protegido. Estos sistemas pueden identificar indicadores de actividades delictivas como tráfico de datos anómalo o patrones de comunicación sospechosos, manteniendo un equilibrio entre seguridad y privacidad con tasas de falsos positivos inferiores al 0.1%.[45]

---

42 Kaur, Parneet, Yogesh Kumar, and Surbhi Gupta. "Artificial intelligence techniques for the recognition of multi-plate multi-vehicle tracking systems: a systematic review." *Archives of Computational Methods in Engineering* 29, no. 7 (2022): 4897-4914.

43 https://news.microsoft.com/es-xl/features/como-photodna-for-video-es-utilizado-para-combatir-la-explotacion-infantil-en-linea/

44 https://www.feedzai.com/resource/how-latam-banks-can-measure-classify-and-prevent-scams/

45 Wu Xu-dong y Li Xin. "Protect Cloud Computing's Data Using Fully Homomorphic Encryption." **National Conference on Information Technology and Computer Science (CITCS 2012).** Publicado por Atlantis Press, 2012. Disponible en: http://eprint.iacr.org/2009/571.

EXPERIENCIAS EN AMÉRICA LATINA

La implementación de sistemas de IA en seguridad pública en América Latina presenta casos significativos que ilustran tanto los alcances como las limitaciones de estas tecnologías. El caso de Uruguay (2013-2017) con *Predpol* resulta particularmente ilustrativo. Este software de predicción delictiva elaboraba mapas detallados para optimizar el despliegue policial, basándose en cuatro variables principales: tipo de delito, lugar, fecha y hora de comisión. La evaluación del sistema en Montevideo, donde la mitad de las seccionales utilizaron Predpol y la otra mitad empleó métodos tradicionales de análisis estadístico, no encontró diferencias significativas en la efectividad, lo que llevó a la suspensión del programa. Sin embargo, el país continuó su apuesta por la tecnología, y en 2020 su Senado aprobó la creación de una base de datos de identificación facial para fines de seguridad pública.[46]

Colombia presenta dos casos emblemáticos. Por una parte, PRISMA el cual está diseñado para predecir el riesgo de reincidencia delictiva, representa un intento de sistematizar las decisiones sobre detención preventiva. El sistema recopila información exhaustiva sobre individuos investigados, incluyendo arrestos previos, procesos en trámite y antecedentes de reclusión. Sin embargo, su implementación ha generado preocupaciones similares a las observadas en Estados Unidos, donde sistemas análogos han mostrado sesgos raciales en sus predicciones.[47]

Por otra parte, el Centro de Comando, Control, Comunicaciones y Cómputo (C4) de Bogotá representa otro caso notable. Este sistema integra 6,000 cámaras de videovigilancia con capacidades de reconocimiento facial y análisis de voz, permitiendo el rastreo de delincuentes mediante el filtrado de características específicas en tiempo real. El proyecto busca expandirse mediante un acuerdo con la Registraduría Nacional para acceder a datos biométricos de identificación nacional, aunque esto ha generado preocupaciones sobre privacidad y potencial discriminación.

En Argentina, la implementación en 2019 de un sistema de reconocimiento facial en Buenos Aires, integrado con la infraestructura de cámaras de seguri-

46 OECD/CAF. *Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe.* Estudios de la OCDE sobre Gobernanza Pública. Paris: OECD Publishing, 2022. https://doi.org/10.1787/5b189cb4-es., 47.

47 Partnership on AI. "Artificial Intelligence Research and Ethics Community Calls for Standards in Criminal Justice Risk Assessment Tools." Publicado el 26 de abril de 2019 en https://partnershiponai.org/resource/artificial-intelligence-research-and-ethics-community-calls-for-standards-in-criminal-justice-risk-assessment-tools/

dad existente, marcó un hito en la región. La posterior aprobación de una ley local que autoriza excepciones para el uso de estas tecnologías en materia de seguridad pública estableció un precedente legislativo importante.[48]

Ecuador desarrolló el ambicioso sistema ECU 911, una red nacional que integra 4,300 cámaras de vigilancia y 16 centros regionales, con más de 3,000 empleados monitoreando imágenes y respondiendo llamadas. El sistema incorpora tecnologías avanzadas como cámaras térmicas para monitoreo volcánico, drones de visión nocturna y una plataforma automatizada para envío de evidencias a tribunales.[49] Sin embargo, ha enfrentado críticas significativas, particularmente después de que una investigación de The New York Times revelara que los videos se compartían con la agencia de seguridad nacional del país.[50]

Brasil ha tomado un enfoque más focalizado, como lo demuestra la ley aprobada en Brasilia en noviembre de 2020 que autoriza el reconocimiento facial para seguridad pública. La justificación se basó en la efectividad demostrada durante el carnaval de ese año en la reducción de índices delictivos.[51]

MARCO REGULATORIO Y DESAFÍOS JURÍDICOS

El panorama regulatorio en América Latina es heterogéneo y refleja las diferentes aproximaciones que los países de la región han adoptado frente al desafío de regular la IA en el ámbito de la seguridad pública. Perú marcó un hito al aprobar la primera ley sobre IA en la región a mediados de 2023, aunque con alcances limitados que intentaron ser subsanados mediante un reglamento que aún presenta deficiencias técnicas significativas para su implementación efectiva.[52]

48 Asociación por los Derechos Civiles. "#ConMiCaraNo: Reconocimiento facial en la Ciudad de Buenos Aires." Publicado el 23 de mayo de 2019. https://adc.org.ar/2019/05/23/con-mi-cara-no-reconocimiento-facial-en-la-ciudad-de-buenos-aires/

49 OECD/CAF. *Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe.* Estudios de la OCDE sobre Gobernanza Pública. Paris: OECD Publishing, 2022. https://doi.org/10.1787/5b189cb4-es., 49

50 Mozur, Paul, Jonah M. Kessel y Melissa Chan. "Made in China, Exported to the World: The Surveillance State." *The New York Times*, 24 de abril de 2019. https://www.nytimes.com/es/2019/04/24/espanol/america-latina/ecuador-vigilancia-seguridad-china.html

51 Interlira Reports. "The Implementation of Facial Recognition Cameras for Public Security in Brazil: Benefits and Controversies." Publicado el 4 de mayo de 2022. https://interlira-reports.com/es/news/the-implementation-of-facial-recognition-cameras-for-public-security-in-brazil-benefits-and-controversies/05/04/2022/

52 Lara, J. Carlos. "Inteligencia Artificial en América Latina: Regulación no significa protección." *Derechos Digitales*, 7 de junio de 2024. https://www.derechosdigitales.org/23866/

Argentina, Colombia, Brasil y México mantienen proyectos de ley en discusión, cada uno con aproximaciones distintas que evidencian la diversidad de interpretaciones sobre la IA y sus aplicaciones en seguridad. Destaca que únicamente el proyecto brasileño aborda de manera específica las aplicaciones en seguridad pública, dejando un vacío significativo en las demás propuestas legislativas.[53]

El caso chileno merece especial atención por su particular evolución. Tras un primer intento fallido de regulación en el Congreso, el ejecutivo presentó un proyecto propio fuertemente influenciado por factores económicos y de seguridad pública. La propuesta no solo omite medidas preventivas para contener potenciales abusos en el uso de sistemas de reconocimiento facial automatizado, sino que explícitamente rechaza esta necesidad al otorgar un estatus de excepcionalidad a las esferas de seguridad nacional y pública. Como lo expresó la Ministra de Ciencia de Chile, Aisén Etcheverry, se reconoce que aunque el análisis biométrico en tiempo real en espacios públicos conlleva altos riesgos de vulneración de derechos fundamentales, se justifica su uso en el contexto de la seguridad y la persecución del crimen organizado, siempre que sea realizado por autoridades competentes autorizadas por ley.

La mayoría de los proyectos legislativos en la región siguen el enfoque basado en riesgos establecido por el Acta de Inteligencia Artificial de la Unión Europea de 2023, que clasifica los sistemas de IA según su nivel de riesgo e impone regulaciones más estrictas a las aplicaciones consideradas de alto riesgo. Sin embargo, existe una tendencia preocupante hacia la implementación de sistemas de vigilancia biométrica en espacios públicos sin la transparencia necesaria, práctica que se observa en Argentina, Brasil, Ecuador, Uruguay, Colombia, Bolivia, Paraguay y México.

Esta tendencia contrasta marcadamente con las restricciones implementadas en otras partes del mundo. En Estados Unidos, ciudades como San Francisco, Somerville y Oakland han prohibido el uso de reconocimiento facial por parte de las fuerzas de seguridad pública. Incluso empresas tecnológicas líderes como IBM, Amazon y Microsoft han solicitado una regulación más estricta, reconociendo los riesgos inherentes a estas tecnologías.

inteligencia-artificial-en-america-latina-regulacion-no-significa-proteccion/

53 Pablo Contreras, "Convergencia Internacional y Caminos Propios: Regulación de la Inteligencia Artificial en América Latina," *Actualidad Jurídica Iberoamericana* 21 (agosto 2024): 468-493.

RIESGOS Y PREOCUPACIONES

La implementación de IA en seguridad pública enfrenta desafíos significativos que van más allá de las consideraciones técnicas. Un informe de Fair Trials de 2024 sobre siete países latinoamericanos documenta vulneraciones a derechos fundamentales como la privacidad, la igualdad, la no discriminación, la presunción de inocencia y el debido proceso. Estas preocupaciones se magnifican por la falta de transparencia gubernamental en las iniciativas que integran estos sistemas en los sistemas penales.

Un fenómeno particular que caracteriza la implementación de IA en la región es el "solucionismo tecnológico", la tendencia a intentar resolver problemas estructurales mediante la implementación de herramientas tecnológicas sin abordar las causas subyacentes. Este enfoque ha resultado particularmente problemático en el contexto latinoamericano, donde las desigualdades sociales profundas requieren aproximaciones más holísticas. Paradójicamente, una de las áreas menos exploradas desde el punto de vista de política pública es la incorporación de soluciones de IA para el estudio de factores sociales como pobreza, educación y salud desde la perspectiva de seguridad pública.

Los sistemas de reconocimiento facial presentan preocupaciones específicas que merecen atención detallada. Los expertos distinguen entre diferentes técnicas: la autenticación (comparación uno a uno, como en control de accesos), la identificación (comparación uno a muchos, como en búsqueda de sospechosos), y la clasificación (análisis de rostros para categorización según criterios como género o etnia). Cada una de estas técnicas presenta desafíos éticos y técnicos particulares.

Los sistemas de reconocimiento facial presentan preocupaciones específicas que merecen atención detallada. Los expertos distinguen entre diferentes técnicas, cada una con sus propios desafíos éticos y técnicos:

La autenticación (comparación uno a uno) se utiliza principalmente en controles de acceso y verificación de identidad. En términos técnicos, estos sistemas enfrentan desafíos con la calidad y consistencia de la captura de imágenes, variaciones en iluminación y el envejecimiento natural de los rostros. Los desafíos éticos incluyen el almacenamiento y protección de datos biométricos, el consentimiento informado para su recolección, y el derecho al olvido. Por ejemplo, cuando un empleado deja una organización, surge la pregunta sobre la retención o eliminación de sus datos biométricos.

La identificación (comparación uno a muchos) presenta desafíos técnicos más complejos debido a la necesidad de comparar un rostro contra bases de datos extensas. Los sistemas actuales muestran tasas de falsos positivos

significativamente más altas cuando operan a esta escala. Un estudio del NIST reveló que los algoritmos pueden tener tasas de error hasta 100 veces mayores en ciertos grupos demográficos, particularmente en personas de color y mujeres. Los desafíos éticos son especialmente preocupantes: el uso de estas tecnologías para vigilancia masiva, la potencial discriminación racial y étnica, y el riesgo de identificaciones erróneas que pueden llevar a detenciones injustificadas.

La clasificación, que analiza rostros para categorización según criterios como género o etnia, presenta los desafíos éticos más controversiales. Técnicamente, estos sistemas pueden perpetuar y amplificar sesgos existentes, especialmente cuando se entrenan con conjuntos de datos no representativos de la diversidad racial y étnica de América Latina. La precisión de estos sistemas cae significativamente al analizar rostros de grupos subrepresentados en los datos de entrenamiento. Los desafíos éticos incluyen el refuerzo de estereotipos, la legitimación de perfiles raciales, y el uso potencial para discriminación sistemática.

HACIA UNA IMPLEMENTACIÓN RESPONSABLE

Para garantizar una implementación responsable de la IA en seguridad pública, es fundamental establecer un marco de principios rectores robustos y aplicables. La equidad debe constituir un pilar fundamental, asegurando que los sistemas respeten derechos básicos como el debido proceso, la presunción de inocencia y la libertad de expresión. Esto implica no solo el diseño técnico de los sistemas, sino también su implementación y monitoreo continuo.

La rendición de cuentas debe institucionalizarse a nivel organizacional, estableciendo mecanismos claros de supervisión y responsabilidad. Esto requiere la creación de estructuras de gobierno específicas para la gestión de sistemas de IA, con roles y responsabilidades claramente definidos.

La transparencia emerge como un principio crítico, especialmente en el contexto de seguridad pública. Los procesos de toma de decisiones de los sistemas no deben ser "cajas negras", sino que deben permitir auditoría y escrutinio. Esto es particularmente relevante cuando las decisiones afectan directamente los derechos y libertades de los ciudadanos.

La "explicabilidad" de los sistemas debe garantizar que las acciones y decisiones sean comprensibles no solo para los usuarios técnicos, sino también para los operadores del sistema judicial y los ciudadanos afectados. Esto implica desarrollar interfaces y documentación que permitan entender el razonamiento detrás de las decisiones automatizadas.

Considerando lo anterior, para una implementación efectiva, se recomienda:

1. Establecer marcos de evaluación de impacto en derechos humanos antes de la implementación de sistemas de IA en seguridad pública.
2. Desarrollar protocolos específicos para el uso de tecnologías de reconocimiento facial y biométrico.
3. Implementar sistemas de auditoría regular que evalúen el desempeño y los sesgos potenciales.
4. Crear mecanismos de participación ciudadana en la supervisión y evaluación de estos sistemas.
5. Establecer límites claros en el uso de datos personales y biométricos.

INTELIGENCIA ARTIFICIAL Y EL SOLUCIONISMO TECNOLÓGICO: CONSIDERACIONES FINALES

La integración de la IA en la seguridad pública en América Latina representa una encrucijada crítica entre la innovación tecnológica y la protección de derechos fundamentales. Si bien estas tecnologías ofrecen oportunidades significativas para mejorar la eficiencia y efectividad de las fuerzas de seguridad, su implementación debe equilibrarse cuidadosamente con salvaguardas robustas que protejan los derechos individuales y prevengan abusos.

La experiencia acumulada en la región demuestra tanto el potencial como los riesgos inherentes a estas tecnologías. Casos como el de Predpol en Uruguay ilustran que la implementación de soluciones tecnológicas no siempre produce los resultados esperados, mientras que experiencias como el ECU 911 en Ecuador y el C4 en Colombia evidencian la necesidad de mayor transparencia y supervisión en el manejo de datos sensibles.

La región enfrenta el desafío particular de desarrollar marcos regulatorios que permitan aprovechar los beneficios de la IA mientras se protegen los derechos fundamentales en un contexto de pronunciadas desigualdades sociales. El éxito en esta tarea dependerá de la capacidad de los gobiernos para establecer políticas pública de Estado que promuevan la transparencia, la rendición de cuentas y la participación ciudadana en el desarrollo y despliegue de estas tecnologías.

Es fundamental recordar que cualquier injerencia en los derechos individuales mediante el uso de IA debe estar debidamente sustentada, ser legal, necesaria y proporcional. Esto requiere no solo marcos regulatorios adecuados, sino también mecanismos efectivos de supervisión y la voluntad política de implementarlos.

Un fenómeno particular que caracteriza la implementación de IA en la región es el "solucionismo tecnológico", la tendencia a intentar resolver problemas estructurales mediante la implementación de herramientas tecnológicas sin abordar las causas subyacentes. Este enfoque, acuñado por Evgeny Morozov, se manifiesta cuando las autoridades priorizan soluciones tecnológicas inmediatas sobre reformas estructurales de largo plazo. Por ejemplo, mientras ciudades como Río de Janeiro han invertido millones en sistemas de videovigilancia con IA para monitorear zonas de alto riesgo, los programas de prevención social y desarrollo comunitario en estas mismas áreas reciben una fracción del presupuesto.

El solucionismo tecnológico en seguridad pública se evidencia en tres patrones principales: Primero, la implementación de sistemas de vigilancia masiva sin un marco regulatorio adecuado. Por ejemplo, el despliegue de cámaras con reconocimiento facial en Buenos Aires se realizó antes de contar con protocolos completos sobre el manejo de datos biométricos.

El segundo patrón trata sobre la adopción de tecnologías predictivas sin considerar sesgos sociales preexistentes. El sistema CompStat 2.0, implementado en varias ciudades latinoamericanas, ha sido criticado por perpetuar patrones de vigilancia excesiva en comunidades marginadas, basándose en datos históricos que reflejan sesgos sistemáticos en la aplicación de la ley.

El tercer patrón es la falta de integración entre sistemas tecnológicos y programas sociales. Mientras se invierten recursos significativos en herramientas de IA para detectar patrones criminales, existe un vacío en el uso de estas mismas tecnologías para identificar y abordar factores de riesgo social.

Este enfoque ha resultado particularmente problemático en el contexto latinoamericano, donde las desigualdades sociales profundas requieren aproximaciones más holísticas. La región presenta los índices de desigualdad más altos del mundo. Esta realidad se refleja en patrones de criminalidad estrechamente vinculados a factores socioeconómicos.

Paradójicamente, una de las áreas menos exploradas desde el punto de vista de política pública es la incorporación de soluciones de IA para el estudio de factores sociales como pobreza, educación y salud desde la perspectiva de seguridad pública. Existen ejemplos prometedores a nivel global que podrían adaptarse al contexto latinoamericano. Por ejemplo, el programa *Risk Terrain Modeling*, implementado en ciudades como Newark y Minneapolis, utiliza IA para analizar no solo patrones criminales, sino también factores ambientales y socioeconómicos que contribuyen a la inseguridad. Este enfoque ha permitido reducir el crimen en áreas específicas hasta en un 35% mediante intervenciones

focalizadas que combinan vigilancia con mejoras en infraestructura y programas sociales.

Algunas iniciativas emergentes en la región comienzan a abordar esta brecha. Un proyecto llevado a cabo por la Universidad de la Amazonia está desarrollando un sistema de IA que integra datos de criminalidad con indicadores sociales como deserción escolar, desempleo juvenil y acceso a servicios básicos. Este enfoque permite identificar áreas de intervención prioritaria donde las mejoras en condiciones sociales podrían tener el mayor impacto en la reducción del crimen.

La superación del solucionismo tecnológico requiere un cambio de paradigma en la implementación de IA en seguridad pública. Es necesario desarrollar marcos de evaluación que consideren no solo la eficacia técnica de las soluciones, sino también su impacto social y su capacidad para abordar causas estructurales. Esto implica la creación de equipos multidisciplinarios que incluyan no solo expertos en tecnología y seguridad, sino también especialistas en desarrollo social, antropólogos y representantes comunitarios.

IV. CONCLUSIÓN Y HALLAZGOS

El análisis desarrollado a lo largo de este capítulo revela patrones significativos en la implementación de la inteligencia artificial para abordar la desinformación y la seguridad pública en América Latina y el Caribe. Lejos de ofrecer soluciones unívocas, la evidencia sugiere que el éxito de estas implementaciones depende crucialmente de la capacidad institucional, el marco regulatorio y el contexto sociopolítico de cada país.

Una lección fundamental emerge del estudio: la necesidad de superar el falso dilema entre innovación tecnológica y protección de derechos fundamentales. Las experiencias más exitosas en la región demuestran que estos objetivos no solo son compatibles, sino mutuamente reforzantes. Los casos de implementación analizados indican que cuando la tecnología se despliega con salvaguardas adecuadas y supervisión democrática, puede fortalecer simultáneamente la seguridad ciudadana y la confianza en las instituciones.

El panorama regional señalado durante el capítulo también revela una paradoja instructiva: mientras algunos países han logrado avances significativos en la implementación técnica de sistemas de IA, frecuentemente han descuidado aspectos cruciales como la participación ciudadana y la transparencia. Esta disparidad ha resultado en una brecha creciente entre la sofisticación tecnológica y la legitimidad social de estas herramientas, subrayando la importancia de desarrollar aproximaciones más integrales.

Un hallazgo particularmente relevante es la interconexión entre los desafíos de desinformación y seguridad pública. La experiencia latinoamericana demuestra que estos fenómenos no operan en aislamiento: la desinformación puede exacerbar crisis de seguridad, mientras que la inseguridad puede crear condiciones propicias para la proliferación de narrativas falsas. Esta interrelación sugiere la necesidad de estrategias coordinadas que aborden ambas problemáticas de manera coherente.

El análisis realizado durante el presente capítulo también revela una tendencia preocupante: la tentación de importar soluciones tecnológicas sin adaptarlas adecuadamente al contexto regional. La efectividad limitada de algunos sistemas predictivos y de reconocimiento facial sugiere que las soluciones desarrolladas para otros contextos no necesariamente responden a las realidades sociales, culturales y económicas de América Latina. Esta observación resalta la importancia de desarrollar capacidades locales de innovación y adaptación tecnológica.

Mirando hacia el futuro, emergen tres imperativos fundamentales para la región. Primero, la necesidad de desarrollar marcos regulatorios que promuevan la innovación mientras protegen efectivamente los derechos ciudadanos. Segundo, la importancia de fortalecer las capacidades institucionales para implementar y supervisar sistemas de IA. Tercero, la urgencia de fomentar un ecosistema de innovación que responda a las necesidades específicas de la región.

La evidencia presentada también sugiere la necesidad de replantear el papel de la sociedad civil, la academia y el sector privado en estos procesos. Las experiencias más exitosas han involucrado activamente a estos actores no solo en la supervisión, sino también en el diseño e implementación de soluciones. Este enfoque participativo no solo mejora la legitimidad de las intervenciones tecnológicas, sino que también contribuye a su efectividad y sostenibilidad.

Finalmente, el análisis subraya que la implementación de la IA en estos ámbitos críticos no puede desvincularse de objetivos más amplios de desarrollo social y fortalecimiento democrático. La tecnología, por más sofisticada que sea, debe concebirse como una herramienta al servicio de estos objetivos fundamentales, no como un fin en sí misma. Sólo a través de esta comprensión más amplia podrá América Latina aprovechar verdaderamente el potencial transformador de la IA para construir sociedades más seguras, informadas y democráticas.

V. BIBLIOGRAFÍA

ACTIVA & Worlwide Independent Network of Market Research. 2022. "Fake news: Desinformación en Chile y LatAm." https://chile.activasite.com/wp-content/uploads/2023/04/222287_Estudio-Desinformacio%CC%81n_0329_V1-2.pdf

Alketbi, Salem K. 2024. "Emerging Technologies in Forensic DNA Analysis." Available at SSRN 4956325.

Alonso-González, Marián & Hada M. Sánchez Gonzales. 2024. "Inteligencia artificial en la verificación de la información política. Herramientas y tipología." Revista Más Poder Local 56: 30.

Arreola García, Adolfo. 2024. "Inteligencia Artificial y Desinformación: Papel de los Conflictos del Siglo XXI." Centro de Estudios Estratégicos del Ejército de Perú 3 (3): 87-113.

Asociación por los Derechos Civiles. 2019. "#ConMiCaraNo: Reconocimiento facial en la Ciudad de Buenos Aires." https://adc.org.ar/2019/05/23/con-mi-cara-no-reconocimiento-facial-en-la-ciudad-de-buenos-aires/

Atanasova, Pepa, Jakob Grue Simonsen, Christina Lioma & Isabelle Augenstein. 2020. "Generating fact checking explanations." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 7352-7364.

Ball, Joshua. 2022. "The changing face of conflict: What is hybrid warfare?" Global Security Review. https://globalsecurityreview.com/hybrid-and-non-linear-warfare-systematically-erases-the-divide-between-war-peace/

Barredo-Ibáñez, Daniel, Daniel Javier De la Garza Montemayor, Ángel Torres Toukoumidis & Paulo Carlos López-López. 2021. "Artificial intelligence, communication, and democracy in Latin America: a review of the cases of Colombia, Ecuador, and Mexico." Profesional de la Información 30 (6): 1-17.

Bennet, Lance & Steven Livingston. 2018. "The disinformation order: Disruptive communication and the decline of democratic institutions." European Journal of Communication 32 (2): 122-139.

Bradshaw, Samantha & Philip N. Howard. 2017. Troops, trolls and troublemakers: a global inventory of organized social media manipulation. Oxford: University of Oxford.

Broda, Elena & Jesper Strömbäck. 2024. "Misinformation, disinformation, and fake news: lessons from an interdisciplinary, systematic literature review." Annals of the International Communication Association 48 (2): 139-166.

Campos-Domínguez, Eva & Berta García-Orosa. 2018. "Comunicación algorítmica en los partidos políticos: automatización de producción y circulación de mensajes." Profesional de la Información 27 (4): 769-777.

Cortina, Adela. 2004. "Ciudadanía mediática." El País. https://elpais.com/diario/2004/11/24/opinion/1101250805_850215.html

Demartini, Gianluca, Stefano Mizzaro, & Damiano Spina. 2020. "Human-in-the-loop Artificial Intelligence for Fighting Online Misinformation: Challenges and Opportunities." IEEE: 65-75.

Diakopoulos, Nicholas & Deborah Johnson. 2020. "Anticipating and addressing the ethical implications of deepfakes in the context of elections." New Media & Society 23 (7): 2072-2098.

Duchiade, André. 2023. "Nuevo mapa destaca leyes contra la desinformación en América Latina y el Caribe." LatAm Journalism Review. https://latamjournalismreview.org/es/articles/nuevo-mapa-destaca-leyes-contra-la-desinformacion-en-america-latina-y-mas-alla/

Durnová, Anna. 2019. "Unpacking emotional contexts of post-truth." Critical Policy Studies 13 (4): 447-450.

Endert, Julius. 2024. "Generative AI is the ultimate disinformation amplifier." Deutsche Welle. https://akademie.dw.com/en/generative-ai-is-the-ultimate-disinformation-amplifier/a-68593890

French, Aaron M., Veda C. Storey & Linda Wallace. 2023. "The impact of cognitive biases on the believability of fake news." European Journal of Information Systems: 1-22.

Interpol and United Nations Interregional Crime and Justice Research Institute (UNICRI). 2019. "Artificial Intelligence and Robotics for Law Enforcement." https://unicri.it/sites/default/files/2019-11/Artificial%20Intelligence%20and%20Robotics%20for%20Law%20Enforcement.pdf

Interlira Reports. 2022. "The Implementation of Facial Recognition Cameras for Public Security in Brazil: Benefits and Controversies." https://interlira-reports.com/es/news/the-implementation-of-facial-recognition-cameras-for-public-security-in-brazil-benefits-and-controversies/05/04/2022/

Kaur, Parneet, Yogesh Kumar, and Surbhi Gupta. 2022. "Artificial intelligence techniques for the recognition of multi-plate multi-vehicle tracking systems: a systematic review." Archives of Computational Methods in Engineering 29 (7): 4897-4914.

Kennan, George F. 1948. "The inauguration of organized political warfare." Wilson Center. https://digitalarchive.wilsoncenter.org/document/george-f-kennan-inauguration-organized-political-warfare

Labuz, Mateusz & Christopher Nehring. 2024. "On the way to deep fake democracy? Deep fakes in election campaigns in 2023." European Political Science. https://doi.org/10.1057/s41304-024-00482-9

Lara, J. Carlos. 2024. "Inteligencia Artificial en América Latina: Regulación no significa protección." Derechos Digitales. https://www.derechosdigitales.org/23866/inteligencia-artificial-en-america-latina-regulacion-no-significa-proteccion/

Li, Cathy & Agustina Callegari. 2024. "Stopping AI disinformation: Protecting truth in the digital world." World Economic Forum. https://www.weforum.org/agenda/2024/06/ai-combat-online-misinformation-disinformation/

López-López, Paulo Carlos, Nieves Lagares Díez & Iván Puentes-Rivera. 2021. "La inteligencia artificial contra la desinformación: una visión desde la comunicación política." Razón y Palabra 24 (111): 5-11.

Manfredi Sánchez, Juan Luis & María José Ufarte Ruiz. 2020. "Inteligencia artificial y periodismo: una herramienta contra la desinformación." Revista CIDOB d'Afers Internacionals 124: 49-72.

Moreno Espinosa, Pastora, Abdulsalam Alsarayreh & Juan Carlos Figuereo Benítez. 2024. "El big data y la inteligencia artificial como soluciones a la desinformación." Doxa Comunicación 38.

Mozur, Paul, Jonah M. Kessel, and Melissa Chan. 2019. "Made in China, Exported to the World: The Surveillance State." The New York Times, April 24, 2019. https://www.nytimes.com/es/2019/04/24/espanol/america-latina/ecuador-vigilancia-seguridad-china.html

OECD/CAF. 2022. Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe. Estudios de la OCDE sobre Gobernanza Pública. Paris: OECD Publishing. https://doi.org/10.1787/5b189cb4-es

O'Sullivan, Shane, Andreas Holzinger, Kurt Zatloukal, Paulo Saldiva, Mohammed Imran Sajid, and Dominic Wichmann. 2017. "Machine learning enhanced virtual autopsy." Autopsy & case reports 7 (4): 3.

Partnership on AI. 2019. "Artificial Intelligence Research and Ethics Community Calls for Standards in Criminal Justice Risk Assessment Tools." https://partnershiponai.org/resource/artificial-intelligence-research-and-ethics-community-calls-for-standards-in-criminal-justice-risk-assessment-tools/

Rosenzvit, Milena. 2022. "Lanzamos la Red latinoamericana de formadores en fact checking." Chequeado. https://chequeado.com/lanzamos-la-red-latinoamericana-de-formadores-en-fact-checking/

Sacco, M. A., P. Tarzia, L. Tarda, R. La Russa, F. Cordasco, and I. Aquila. 2024. "The artificial intelligence in autopsy and crime scene analysis." La Clinica Terapeutica 175 (4).

Sánchez Gonzales, Hada. 2022. "Transformación digital y audiencia. Tendencia y uso digital en medios verificadores." Ámbitos. Revista Internacional de Comunicación 56: 9-20.

Serra-García, Marta & Uri Gneezy. 2023. "Improving human deception detection using algorithmic feedback." CESIFO Working Papers 10518. https://www.cesifo.org/en/publications/2023/working-paper/improving-human-deception-detection-using-algorithmic-feedback

Snyder, Timothy. 2017. On Tiranny: Twenty Lessons from the Twentieth Century. New York: Crown Publishing.

Soprano, Michael, Kevin Roitero, David La Barbera, David Ceolin, Damiano Spina, Gianluca Demartini & Stefano Mizzaro. 2024. "Cognitive Biases in Fact-Checking and Their Countermeasures: A Review." Information Processing and Managment 61: 1-29.

Stears, Chris, and Joshua Deeks. 2023. "The use of artificial intelligence in fighting financial crime, for better or worse?" Journal of Money Laundering Control 26 (3): 433-435.

Šusteková, D., and Marta Knutelska. 2015. «How is the artificial intelligence used in applications for traffic management.» Machines. Technologies. Materials. 9 (10): 49-52.

WEF. 2024. "Global Risks Report 2024." World Economic Forum. https://www.weforum.org/publications/global-risks-report-2024/

Wu, Xu-dong, and Li Xin. 2012. "Protect Cloud Computing's Data Using Fully Homomorphic Encryption." In National Conference on Information Technology and Computer Science (CITCS 2012). Atlantis Press. http://eprint.iacr.org/2009/571

AUTORES

> ALAM IQBAL, BADAR

Investigador en el Departamento de Economía, Universidad de Stellenbosch, Sudáfrica. Es un reconocido académico con más de cuatro décadas de experiencia en los campos de la economía internacional y los negocios. Inició su carrera en 1978 como Profesor Asistente en el Departamento de Comercio de la Universidad Musulmana de Aligarh (AMU), India, donde ascendió a Profesor Asociado en 1986 y a Profesor Titular en 1998. Ocupó el cargo de Decano de la Facultad de Comercio de la AMU entre 2006 y 2008, y fue Rector Interino en dos ocasiones. Además, ha sido miembro de la Corte Universitaria, el Consejo Ejecutivo y el Consejo Académico de la AMU. A lo largo de su trayectoria, ha trabajado como Profesor Visitante en diversas instituciones internacionales, como Claflin University y Kentucky State University en EE.UU. (como Fulbright Scholar-in-Residence), la Universidad de Rennes 1 en Francia, el Instituto de Economías en Desarrollo en Tokio, la Universidad de Sudáfrica, la Universidad Vaal de Tecnología en Sudáfrica, la Universidad de Maringá en Brasil, el Instituto de Tecnología de Beijing en China y la Universidad de Varsovia en Polonia, entre otras. Ha sido becario DAAD en Alemania en tres ocasiones, becario de la Fundación Ford en EE.UU. y profesor Fulbright en dos periodos. Además, ha trabajado como Profesor Adjunto de Economía y Finanzas en la Monarch Business School de Suiza. Sus contribuciones académicas incluyen la publicación de 12 libros y más de 138 artículos en revistas internacionales en países como Suiza, Canadá, Reino Unido, México y Brasil, además de 81 artículos en publicaciones reconocidas en India. Ha participado en conferencias y coloquios internacionales, con especial énfasis en temas relacionados con los países BRICS. Sus áreas de investigación incluyen la economía internacional, los negocios internacionales y la economía del desarrollo.

> BAIGORI, CARLOS

Presidente y CEO de la Agencia Nacional de Telecomunicaciones de Brasil (ANATEL). Licenciado en Ciencias Económicas por la Universidad de Brasília (UnB), con una Maestría y un Doctorado en Economía por la Universidad Católica de Brasília (UCB). Activo en el área de telecomunicaciones desde 2006, fue consultor y analista de información del sector en una asociación de operadores móviles. También se ha desempeñado como profesor de licenciatura en la Universidad Católica de Brasília (UCB), en el Instituto de Educación Superior de Brasília (IESB) y en el Instituto Brasileño de Mercado de Capitales (IBMEC). Además, ha escrito artículos y capítulos de libros publicados en Brasil y en el extranjero. Es funcionario de carrera en Anatel desde 2009, donde previamente ha ocupado los cargos de Superintendente Ejecutivo, Superintendente de Competencia, Superintendente de Control de Obligaciones y Jefe de Asistencia Técnica.

> BERASALUCE IZA, JULEN

Profesor-investigador del Centro de Estudios Económicos de El Colegio de México (COLMEX) Consejero del Instituto Para el Desarrollo Industrial y la Transformación Digital (INADI). Licenciado en Economía por la Universidad del País Vasco. Erasmus Mundus Máster en Economía Cuantitativa por la Université Paris-1 Panthéon-Sorbonne y la Universitat Autònoma de Barcelona y Doctor en Economía por la Universitat Autònoma de Barcelona. Actualmente es profesor investigador del Centro de Estudios Económicos de El Colegio de México y pertenece al Sistema Nacional de Investigadores. Está especializado en temas de política industrial y experiencias de desarrollo en países asiáticos. Autor y coordinador de diversas obras entre las que destacan: Hacia una nación digital. La importancia de la sostenibilidad y la tecnología 5G en el desarrollo industrial. De la revolución industrial a la revolución digital. Hacia una agenda digital para México. Estado desarrollador. Casos exitosos y lecciones para México. Corea y México. Dos estrategias de crecimiento con resultados dispares.

> CARBAJAL GLASS, FAUSTO

Senior Policy Advisor en Miranda Partners, así como socio fundador de Delphi Solutions and Strategic Services, una empresa de consultoría estratégica e investigación científica enfocada en el riesgo político y las dinámicas de seguridad en América Latina. Con experiencia previa en el sector público en México, específicamente en las Secretarías de Relaciones Exteriores y de Gobernación. Internacionalista por la Universidad Iberoamericana, Ciudad de

México, y Maestro en Estudios de Guerra por el King's College London. Graduado de los cursos "Estrategia y Política de Defensa", por el Centro de Estudios Hemisféricos de Defensa William J. Perry, y "Terrorismo y Contraterrorismo", por la Universidad de Leiden, Países Bajos. Actualmente estudia el Doctorado en Desarrollo y Seguridad Nacional, en el Instituto Mexicano de Estudios Estratégicos en Seguridad y Defensa Nacionales (IMEESDN). Es fellow del International Strategy Forum, iniciativa organizada el Fondo Eric & Wendy Schmidt para la Innovación Estratégica. Asimismo, es asociado y senior fellow del Consejo Mexicano de Asuntos Internacionales (COMEXI), senior research fellow del Centro de Investigación para la Paz México (CIPMEX), así como miembro del Strategic Hub for Organised Crime Research (SHOC) del Royal United Services Institute (RUSI), del European Consortium for Political Research (ECPR), de la Red de Norteamericanistas del Centro de Investigaciones sobre América del Norte – UNAM.

> CASTAÑÓN IBARRA, ROSARIO

Investigadora nacional y ex directora de transferencia de tecnología de la Universidad Nacional Autónoma de México. Investigadora Nacional. Ingeniera Química, Maestra en Ingeniería (Planeación) y Doctora en Administración de Organizaciones; todos sus grados fueron obtenidos en la Universidad Nacional Autónoma de México. Es académica del Instituto de Ciencias Aplicadas y Tecnología de la UNAM; ha impartido cursos relacionados con la gestión de la innovación tecnológica, a nivel maestría y doctorado en la UNAM, así como en otras Universidades. También ha sido profesora en diplomados y cursos cortos organizados por diferentes instituciones a nivel nacional. Su trabajo de investigación se ha enfocado en temas tales como inteligencia tecnológica competitiva, vinculación universidad-industria, transferencia de tecnología, gestión del conocimiento y gestión de la propiedad intelectual. En esta última área fue, de 2007 a 2010, la coordinadora de la RED Iberoamericana para el Fortalecimiento de la Cultura de la Propiedad Intelectual. De agosto de 2008 a febrero de 2012 fue Directora de Transferencia de Tecnología de la Universidad Nacional Autónoma de México.

> COELLO COELLO, CARLOS

Miembro del Colegio Nacional. Investigador en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN). Profesor Visitante en el Basque Center for Applied Mathematics en España. Ha impartido cursos de computación evolutiva, optimización evolutiva multi-objetivo,

lenguajes de programación y optimización de ingeniería en el CINVESTAV-IPN, en grado de maestría y doctorado. Además, ha impartido cursos en España, Inglaterra, Argentina, Chile, India, Bolivia, Colombia, Eslovenia y Estados Unidos. Ha presentado artículos científicos en importantes congresos especializados en computación evolutiva tales como el IEEE Congress on Evolutionary Computation (CEC), la Genetic and Evolutionary Computation Conference (GECCO) y el Parallel Problem Solving from Nature (PPSN), entre otros. Tiene una amplia experiencia como consultor experto en su campo. Es Miembro del Foundation Advisory Board de The International AIQT Foundation, que busca establecer un centro competitivo a nivel internacional en las áreas de inteligencia artificial y tecnología cuántica. Además, es Scientific Advisor de la compañía Complexica, en Australia, y fue Senior Advisor del Hunan Zixing AI Research Institute, en China (2017-2020), por nombrar algunos ejemplos.

> DE LOURDES VÁZQUEZ, MARÍA

Decana de la Escuela de Derecho de la Universidad San Andrés (UdeSA, Buenos Aires, Argentina). Directora de Maestría de Propiedad Intelectual e Innovación organizado en conjunto por la Organización Mundial de la Propiedad Intelectual (OMPI). Destacada académica en el ámbito del derecho, especializada en propiedad intelectual e innovación. Posee una sólida formación académica, habiendo obtenido el título de Master of Laws (LL.M.) en la Harvard Law School. Su experiencia y conocimientos la han llevado a participar en eventos internacionales de relevancia; por ejemplo, en abril de 2021, fue invitada como "Chief Guest" en la "International E-conference on Anatomy of Intellectual Property Rights and Human Rights: Impact Analytics", organizada por el Center for Excellence in IPR de la ICFAI Law School en Hyderabad, India. Su liderazgo en el ámbito académico y su compromiso con la educación en propiedad intelectual e innovación han contribuido significativamente al desarrollo de programas especializados en la región, fortaleciendo la colaboración entre instituciones académicas y organismos internacionales como la OMPI.

> FIERRO GONZÁLEZ, AMAURY

Especialista en inteligencia y geopolítica. Actualmente es instructor en Rabdan Academy, la más importante institución educativa en temas de defensa y seguridad en Medio Oriente, y director de aprendizaje en Tech Corp. Maestro en inteligencia por King's College London, licenciado en relaciones internacionales por la Universidad de las Américas Puebla, con especialidad en contraterrorismo en el Instituto Internacional de Administración de Galilea, en Israel. Ha sido

capacitador en varias fuerzas armadas y policiales de América Latina, así como analista de inteligencia fronteriza en Estados Unidos, inteligencia militar en Reino Unido, y gerente de inteligencia en el sector privado en México. Es asociado de la Academia Mexicana de Derecho Internacional, el Instituto Nacional de Administración Pública y la Red Docente del Instituto de Estudios Estratégicos de Guatemala.

> GARCÍA MONTES, MARIEL

Investigadora y candidata a doctorado en Ciencia, Tecnología y Sociedad, MIT. Especialista en la intersección entre medios digitales, sociedad y activismo, con más de 15 años de experiencia en el campo. Actualmente, es doctoranda en el Programa de Ciencia, Tecnología y Sociedad del MIT, afiliada al Berkman Klein Center for Internet and Society en Harvard y al Data + Feminism Lab del MIT. Antes de su doctorado, obtuvo un grado en Estudios de Medios Comparados en el MIT y fue asistente de investigación en el Center for Civic Media. Sus intereses incluyen juventudes y medios, alfabetización digital, privacidad y vigilancia, educación cívica y moral, procesos participativos y feminismo. Actualmente, investiga redes de ayuda mutua y plataformas comunitarias en línea. Previamente, trabajó en fortalecimiento de capacidades tecnológicas como Oficial de Programas Junior en la Fundación Wikimedia, consultora para la Open Knowledge Foundation e Internews, mentora en UNICEF y miembro de SocialTIC, donde contribuyó en investigación, formación y comunidad en proyectos como Escuela de Datos. También trabajó en estrategia de comunicación en Nova México y fue pasante en el Youth and Media Lab del Berkman Center. Mariel estudió Filosofía en la UNAM y busca generar impacto en la intersección de cultura, tecnología y filosofía, priorizando la acción sobre la teoría.

> GUILLÉN, DAVID

Gerente de consultoría y Gestión de Proyectos en Cambiotec, A.C. Ingeniero mecánico y maestro en Administración Industrial por la UNAM. Especialista en gestión de la innovación tecnológica y experto en estudios de inteligencia tecnológica competitiva. Es evaluador en la Red de Transferencia de Tecnología de la Alianza del Pacífico. Ha dirigido y colaborado en consultorías para organismos a nivel internacional, tal es el caso del Banco Interamericano de Desarrollo (BID), la Organización Mundial de la Propiedad Intelectual (OMPI) y el Ministerio de Industria y Comercio de Panamá; además de gestionar procesos de valuación de tecnologías y transferencia de tecnología para organizaciones como la Unibe (República Dominicana) y el Cicese en México. Ha desarrollado

estudios de vigilancia tecnológica para empresas como Huawei (ciberseguridad), Rotoplas (saneamiento de agua residual), CNH-Industrial (tractor 4.0), e instituciones educativas, entre las que figuran la UNAM, CICY y IPN. Ha liderado a equipos ganadores del Premio ADIAT a la Innovación 2011, Green Tech 2012, Premio BIT 2014 y el Premio Nacional de Tecnología e Innovación. Es coautor del libro De emprendedor a Empresario (2014) y del Manual de Valuación de Tecnología, de la Red de Oficinas de Transferencia de Tecnología (2019).

> JUÁREZ AGUILAR, BEATRIZ

Doctora en Derecho (UNAM) con Maestría sobre China (El Colegio de México) y un Fellowship en la Universidad de Renmin (China). Ha sido docente en la Universidad de Shanxi (Shanxi, China) y en el Posgrado de Derecho de la UNAM. Tiene más de 15 años de experiencia en investigación multidisciplinaria sobre China como centro normativo y familiarizada con el sistema jurídico-político y académico chino. Ha publicado artículos, capítulos de libros y presentado ponencias a nivel nacional e internacional; prepara el libro Introducción al derecho en China. Cuenta con un nivel de chino intermedio-avanzado (HSK4), permitiéndole analizar fuentes en chino.

> LÓPEZ BLANCO, CARLOS

Patrono de la Fundación Euroamérica. Presidente de la Comisión de Economía Digital de la International Chamber of Commerce. Experto en economía digital y en regulación del mercado de las telecomunicaciones. Ha sido director general de Asuntos Públicos y Regulación de Telefónica del 2014 al 2018. Actualmente es presidente de la Comisión de Digitalización de la Cámara de Comercio de España y consejero de Telefónica Argentina. López Blanco es licenciado en derecho por la universidad de Valladolid y diplomado en alta dirección de empresas por el IESE. Durante 1984 hasta 1989 prestó servicios como abogado del estado en el País Vasco y los Ministerios de Educación, Justicia, Economía y Hacienda. En 1989 pasa al sector empresarial como secretario del Consejo de Administración y director de la Asesoría Jurídica de Construcciones Aeronáuticas, S.A. (CASA), empresa española que formaba parte del Consorcio AIRBUS. En 1992, fue nombrado director de la asesoría jurídica en IBM España. De 1995 a 2001 ocupó la Secretaría General de Airtel, que pasó a convertirse en Vodafone España. Fue secretario de estado de telecomunicaciones de 2002 a 2004 y para la Sociedad de la Información en el Gobierno Español.

> LÓPEZ-PORTILLO ROMANO, JOSÉ RAMÓN

Profesor de Oxford, cofundador del Centro de Estudios Mexicanos y miembro del Consejo Científico Internacional. Académico, diplomático, empresario y servidor público mexicano. Ha sido subsecretario de Estado en México. Representante Permanente y Presidente Independiente del Consejo de la FAO. Experto en el impacto socioeconómico de la ciencia, la tecnología y la Innovación. Ha sido asesor del Gobierno de México en estos temas, impulsando las resoluciones sobre el impacto del cambio tecnológico en la Agenda 2030 en Naciones Unidas. Como miembro del "Grupo de los Diez Expertos sobre el Mecanismo de Facilitación de la Tecnología", ha asesorado al Secretario General de las Naciones Unidas y ha contribuido ampliamente a la Guía y la Hoja de ruta hacia los ODS. Es miembro del Consejo Científico Internacional. Recientemente contribuyó a la Cumbre de seguridad de IA de Bletchley Park y al Panel asesor de expertos. También dirige LobsterTel, una empresa de telecomunicaciones en Estados Unidos y México.

> LOUISE SMITH, HANNAH

Investigadora postdoctoral en el Centro de Estudios Avanzados en Derecho de la Innovación en Biociencias (CeBIL) de la Facultad de Derecho de la Universidad de Copenhague. La Dra. Smith se centra en la regulación de nuevas y emergentes tecnologías, incluyendo la informática médica avanzada, la inteligencia artificial y las tecnologías cuánticas. Posee un DPhil en Estudios Socio-Jurídicos de la Universidad de Oxford, así como un M.St en Investigación Socio-Jurídica, un BCL en Derecho y una licenciatura en Jurisprudencia de la misma institución. Gracias a su formación en estudios socio-jurídicos, Hannah tiene un interés particular en cómo las leyes son influenciadas por las normas sociales, el papel del poder en la creación e implementación del derecho y las oportunidades para aprovechar la naturaleza dinámica del derecho para anticipar y adaptarse a los desafíos regulatorios. Sus cargos anteriores incluyen una beca de investigación en la Universidad de Australia Occidental, una cátedra con estipendio en el Magdalen College de la Universidad de Oxford y una pasantía en la Oficina del Gabinete del Reino Unido. Hannah ha publicado en diversas revistas de prestigio, entre ellas The Law Quarterly Review, European Data Protection Law Review y Computer Law and Security Review.

> MANCINI, MALENA

Abogada por la Universidad de Buenos Aires (UBA), especializada en propiedad intelectual y derecho del entretenimiento. Magister en Propiedad Intelectual e Innovación por la Universidad de San Andrés. Docente en la Facultad de Derecho de la UBA, la Universidad de Ciencias Empresariales y Sociales, la Universidad del Museo Social Argentino y la Universidad de San Andrés. Ha realizado una serie de cursos de posgrado en materia de derechos de autor y derechos conexos, derecho digital y nuevas tecnologías, industrias creativas, aspectos legales de la producción audiovisual, gestión de intangibles, entre otros. Es profesora en la Facultad de Derecho de la UBA, la Universidad de Ciencias Empresariales y Sociales (UCES) y la Universidad del Museo Social Argentino (UMSA). Es miembro del Centro de Propiedad Intelectual e Innovación (CPINN) de la Universidad de San Andrés, donde ha formado parte de proyectos de investigación en materia de propiedad intelectual y participa activamente de las diversas propuestas. Es abogada matriculada ante el Colegio Público de la Abogacía de la Capital Federal (C.P.A.C.F.) y Agente de la Propiedad Industrial matriculada ante el Instituto Nacional de Propiedad Industrial (INPI).

> MINSSEN, TIMO

Profesor de Derecho en la Universidad de Copenhague (UCPH) y Director Fundador del Centro de Estudios Avanzados en Derecho de la Innovación en Biociencias (CeBIL). Investigador Afiliado de LML en la Universidad de Cambridge (Reino Unido) y Afiliado de Investigación Inter-CeBIL en el Centro Petrie-Flom de Políticas de Derecho Sanitario, Biotecnología y Bioética en la Facultad de Derecho de Harvard (EE. UU.). Su investigación, supervisión, docencia y práctica de asesoría a tiempo parcial se centran en la Propiedad Intelectual, el Derecho de la Competencia y la Regulación, con un enfoque especial en nuevas tecnologías, big data e inteligencia artificial en el ámbito de la salud y las ciencias de la vida. Timo posee un título en Derecho de la Universidad de Göttingen (Alemania), así como títulos de posgrado en biotecnología y propiedad intelectual (LL.M., M.I.C.L., LL.Lic. y LL.D.) de las universidades de Lund y Uppsala. Ha sido Profesor Visitante Global en la Universidad Técnica de Múnich (TUM) y Becario de Investigación en las universidades de Cambridge y Oxford, la Facultad de Derecho de Harvard, el Chicago-Kent College of Law y el Instituto Max Planck para la Innovación y la Competencia. Además, ha recibido formación en el sistema judicial alemán y en la Oficina Europea de Patentes. Basado en su experiencia académica y práctica en firmas de abogados, tribunales y empresas emergentes de ciencias de la vida, Timo es miembro

de varios comités internacionales y asesor de la OMS, la OMPI, la Comisión Europea, diversas organizaciones, empresas, gobiernos nacionales y firmas de abogados.

> NEGRETE, JORGE FERNANDO

Presidente del Digital Policy & Law Group. Es abogado por parte de la Universidad Nacional Autónoma de México (UNAM). Fue presidente de la Asociación Mexicana de Derecho a la Información A.C. (AMEDI). Es socio presidente de Digital Policy & Law Group, uno de los grupos consultores y de medios más influyentes en América Latina en el área de políticas públicas, regulación de las Telecomunicaciones y Tecnologías de la información. Ha sido profesor de Teoría del Estado, Derecho Administrativo, Instituciones de Derecho Público y Derecho de la Comunicación en la UNAM, ITAM, de la Universidad La Salle y la Universidad Anáhuac del Sur. Su actividad académica y práctica profesional se desarrolla en el Derecho de la Comunicación, la regulación y la política pública de la tecnología digital, la propiedad intelectual, propiedad industrial y transferencia de tecnología. Ha presentado conferencias en Brasil, España, Alemania, Costa Rica, Colombia, China, Suecia y Argentina, entre otros países. Es curador de los foros sobre tecnología digital: México 5G, Colombia 5G, Brasil 5G, Chile 5G y la frontera de la regulación digital.

> OROPEZA GARCÍA, ARTURO

Doctor en Derecho por la Universidad Nacional Autónoma de México (UNAM). Investigador permanente del Instituto de Investigaciones Jurídicas de la UNAM. Actual Presidente del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI). Miembro del Sistema Nacional de Investigadores (SNI) de México. Miembro fundador del Centro de Estudios China-México (CECHIMEX), y del Programa de Estudios sobre Asia y África de la Universidad Nacional Autónoma de México (PUEAA). Miembro del Comité Editorial de China Global Review. Árbitro del Sistema de Controversias del MERCOSUR designado por Brasil según el Protocolo de Olivos. Ha sido conferencista y profesor invitado en diversas universidades e instituciones académicas alrededor del mundo, entre ellas: la Academia China de Ciencias Sociales de Beijing, China; la Universidad Nacional de Derecho de Bangalore, India; la Universidad de Chuo, Japón; Universidad Soongsil, Corea del Sur; Universidad Autónoma de Lisboa, Portugal; Universidad de Montreal, Canadá; Universidad de Johannesburgo, Sudáfrica; la Secretaría General Iberoamericana (SEGIB) en España; la Organización de Estados Americanos (OEA); Asociación Latinoamericana de Integración

(ALADI), Uruguay; la Comisión Económica para América Latina y el Caribe (CEPAL) Chile; entre otros. Es autor y coordinador de más de 40 libros sobre globalización, especializándose en los temas de China, y Sudeste Asiático, entre sus últimas publicaciones se encuentran: "De la Revolución Industrial a la Revolución Digital", Hacia una nueva era del Desarrollo Industrial Digital para México", "Nearshoring. La oportunidad de un nuevo desarrollo económico y social para México", "El T-MEC en el marco de la confrontación China-Estados Unidos", entre otros.

> PANG, ZHONGYING

Académico en humanidades y ciencias sociales, enseña Relaciones Internacionales Globales, Economía Política Internacional e Historia Económica Global en la Universidad de Sichuan en Chengdu, China. Doctor en Derecho en la Universidad de Pekín. Es un reconocido experto en política internacional y actualmente se desempeña como profesor, supervisor doctoral y presidente del Instituto de Desarrollo Marítimo en la Universidad Oceánica de China. También es profesor asociado en la Escuela de Humanidades y Ciencias Sociales de la Universidad de Tsinghua, así como profesor, supervisor doctoral y director del Instituto Global en el Departamento de Relaciones Internacionales de la Universidad de Nankai. Anteriormente, ocupó diversos cargos, incluyendo asistente y profesor en el Departamento de Economía Internacional de la Universidad Renmin de China, investigador asistente y subdirector de la División de Investigación de Relaciones Económicas Internacionales en el Instituto de Economía Mundial y Política de la Academia China de Ciencias Sociales. Además, fue Segundo Secretario de la Embajada de China en Indonesia y miembro del consejo de organizaciones como la Asociación de Naciones Unidas de China y el Foro de Reforma de China. Sus principales áreas de investigación y docencia incluyen temas globales y gobernanza global, estudios sobre Asia y África, política y economía internacional, así como el análisis de políticas diplomáticas.

> PORSDAM MANN, SEBASTIAN

Profesor Asistente en el el Centro de Estudios Avanzados en Derecho de la Innovación Biomédica. Cuenta con una formación académica en filosofía, neurociencia y ética aplicada (licenciatura y doctorado, Universidad de Cambridge) y una reciente investigación doctoral en derecho internacional de los derechos humanos (DPhil en proceso, Universidad de Oxford), aporta una perspectiva interdisciplinaria a su trabajo. Sus nombramientos postdoctorales incluyen cargos en la Escuela de Medicina de Harvard, la Universidad de Oxford

y la Universidad de Copenhague, con el apoyo de una beca personal de la Fundación Carlsberg. Su investigación se centra en explorar el potencial práctico y las condiciones regulatorias necesarias para que nuevas tecnologías –como blockchain y la inteligencia artificial generativa– impulsen el progreso científico y respalden objetivos normativos, en particular en relación con el derecho humano a la ciencia. Porsdam Mann es coautor/editor de The Right to Science: Then and Now (Cambridge University Press, 2021) y Scientific Freedom: The Heart of the Right to Science (Rowman & Littlefield, 2023). Sus contribuciones académicas han sido publicadas en revistas de prestigio como Nature Machine Intelligence, npj Digital Medicine, NEJM AI, PNAS, Philosophical Transactions of the Royal Society, Human Rights Quarterly, Molecular & Cellular Proteomics y American Journal of Bioethics. Su trabajo tiene un impacto de citación ponderado por campo de 5.55 (SciVal, 2021–2025).

> RUÍZ DURÁN, CLEMENTE

Excoordinador del Posgrado en Economía, UNAM, Consejero del Instituto para el Desarrollo Industrial y la Transformación Digital (INADI). Doctor en Economía por la Universidad de Pittsburgh, Pensilvania en Estados Unidos (1975) con especialidad sobre instituciones del bienestar por la Universidad Autónoma de Baja California (2004); profesor en la Facultad de Economía y catedrático de maestría y doctorado en el Posgrado de la Facultad de Economía de la Universidad Nacional Autónoma de México; Coordinador del Campo de Desarrollo Urbano y Regional del Posgrado de Economía; Investigador del Sistema Nacional de Investigadores del Consejo Nacional de Humanidades Ciencia y Tecnología (CONAHCYT). Dentro de sus actividades como docente a nivel licenciatura, maestría y doctorado, ha impartido cátedra de las materias Economía Mexicana I y II, Investigación y Análisis Económico III y IV (1993), Economía Urbana y Regional (2012 – a la fecha), Seminario de Doctorado en Economía Urbana y Regional (2004 – a la fecha), Seminario de Política Industrial (1998-2003), entre otros diversos cursos y seminarios impartidos desde 1987. Además, ha sido asesor y sinodal de más de 200 tesistas a lo largo de su carrera, desde nivel licenciatura hasta nivel de doctorado.

> SAVAGE, SAIPH

Profesora asistente en el Khoury College of Computer Sciences de Northeastern University, donde dirige el Civic A.I. Lab. Su investigación se centra en el diseño, creación y estudio de tecnologías públicas de inteligencia artificial para empoderar a trabajadores, gobiernos, líderes industriales y ONG. Ha asegurado más de $1.52 millones en financiamiento de instituciones como la National Science Foundation (incluyendo el prestigioso NSF CAREER Award), Google, Meta, Naciones Unidas y la OTAN. Su trabajo ha sido reconocido globalmente: fue nombrada una de las 35 Innovadoras Menores de 35 por el MIT Technology Review y una de las 20 líderes en IA en México por Forbes. UNESCO destacó su investigación como una de las más influyentes en IA a nivel mundial. Ha recibido premios en conferencias de alto impacto en interacción humano-computadora, como ACM CHI, CSCW y The Web Conference. Su trabajo ha sido cubierto por medios internacionales como BBC, The New York Times, The Economist, Vice, Wired y Forbes, y apareció en la portada de WIRED en español. En el ámbito de política pública y liderazgo, asesora a gobiernos y senadores en EE.UU. y América Latina sobre temas clave de IA. Es experta en el Global Partnership on AI (GPAI) de la OCDE y ha sido investigadora en el Center for Democracy and Technology (CDT) y la Federation of American Scientists (FAS). También es copresidenta del International Panel on the Information Environment (IPIE), una iniciativa presentada en la Cumbre del Premio Nobel. La Dra. Savage obtuvo su maestría y doctorado en Ciencias de la Computación en la Universidad de California, Santa Bárbara (UCSB) y estudió Ingeniería en Computación en la UNAM. Anteriormente, trabajó en Intel Labs, Microsoft Bing y Stanford, además de haber colaborado con Carnegie Mellon, la Universidad de Washington y West Virginia University, diseñando tecnología para regiones rurales.

> SCHATAN, CLAUDIA

Especialista en temas económicos. Consejera del Instituto Para el Desarrollo Industrial y la Transformación Digital (INADI). Economista mexicana, M.A. en Economía, Universidad de Cambridge, Inglaterra. B.A. en Economía, Universidad de Chile, Santiago, Chile. Consultora de Comisión Económica para América Latina y el Caribe (CEPAL);Consultora para Foro Consultivo de Ciencia y Tecnología (2015); Investigadora Visitante en el Departamento de Economía de la Universidad Iberoamericana, México (2012-2014); Jefa de la Unidad de Comercio Internacional e Industria CEPAL, Naciones Unidas (2000-2011); Oficial de Asuntos Económicos en CEPAL-México (1989-2000) previamente. Catedrática Investigadora del Centro de Investigación y Docencia Económica

(CIDE), México. Investigadora visitante en varias universidades en Estados Unidos (Harvard Institute of International Development, HIID, Universidad de Harvard; Center for United States-Mexican Studies, Universidad de California en San Diego, entre otras).

> SIMG ATILANO, HUGO

Director de Desarrollo de Negocios para América Latina en MediaTek. Maestro en administración de negocios internacionale e ingeniero en electrónica y comunicaciones por el ITESM Campus Estado de México. Fue profesor de ingeniería en el ITESM CCM y en la industria de Telecomunicaciones. Ha desempeñado actividades comerciales y de Desarrollo de negocios, de igual forma ha colaborado en empresas multinacionales fabricantes y desarrolladores de equipos de medición electrónica, dispositivos y sistemas de comunicación móvil, y actualmente en Sistemas en Chip semiconductores.

> SOLLEIRO REBOLLEDO, JOSÉ LUIS

Investigador titular del Instituto de Ciencias Aplicadas y Tecnología (ICAT) de la UNAM. Su producción como investigador comprende más de 200 artículos, capítulos de libros y ponencias publicadas en memorias de congresos internacionales. Miembro del Sistema Nacional de Investigadores en el nivel. Fue Director General de Vinculación de la Universidad Nacional Autónoma de México (UNAM) entre 2008 y 2012. Ha impartido cursos de licenciatura, maestría, doctorado y educación continua en múltiples instituciones de 17 países. Ha asesorado empresas privadas, públicas, universidades, organismos internacionales y asociaciones empresariales en diversas cuestiones relacionadas con la gestión de la innovación. Coordinó el proyecto sobre Sistema Estatal de Innovación en el Estado de México, auspiciado por el Banco Interamericano de Desarrollo y el Consejo Mexiquense de Ciencia y Tecnología. Fue Presidente de la Asociación LES México, A.C., que es el Capítulo mexicano de la Licensing Society International (Asociación de Ejecutivos en Transferencia de Tecnología). Entre los reconocimientos que ha recibido se encuentran: el Premio Jesús Silva Herzog de Investigación Económica y el Premio Ernest Feder de Investigación. Recibió el Doctorado Honoris Causa del Consejo Iberoamericano en Honor a la Excelencia Educativa. En abril del 2017 le otorgaron la distinción "Trayectoria de Mexicano Destacado" en el marco de la Convocatoria para los Reconocimientos a la Innovación de Talento Mexicano Innovation Match 2016-2017.

> SONG, JIN

Directora del Centro de Investigación en Desarrollo, Instituto de Economía y Política Mundial (IWEP) e investigadora de la Academia China de Ciencias Sociales (CASS). Doctora en Economía, Investigadora Senior y Directora del Centro de Investigación en Desarrollo en el Instituto de Economía Mundial y Política (IWEP) de la Academia China de Ciencias Sociales (CASS). Se unió al IWEP, CASS en 2011. Sus áreas de investigación incluyen la economía del desarrollo y la economía laboral, con un enfoque en el mercado laboral, la desigualdad de ingresos y la transición de la estructura industrial. Ha publicado numerosos artículos en revistas especializadas como China Agricultural Economic Review, Journal of World Economy, Journal of Quantitative & Technical Economics, Management World y Studies in Labor Economics. Asimismo, ha contribuido a varios libros publicados por Oxford University Press, Cambridge Press y World Scientific. Es consultora económica del Banco Mundial, colaborando en proyectos de investigación de gran relevancia, incluyendo China 2030, Urbanización en China y Opciones para el Cuidado de Personas Mayores en China.

> SUDOH, OSAMU

Profesor en la Universidad Chuo y profesor en la Escuela de Posgrado en Estudios Interdisciplinarios de la Información (GSII) de la Universidad de Tokio. Obtuvo su doctorado en Economía en la misma institución. A lo largo de su carrera, ha ocupado importantes cargos académicos y de investigación, incluyendo su nombramiento como Profesor en la Escuela de Economía de Estocolmo, Suecia (1995), Profesor Investigador en el Instituto de Investigación de NTT (1997-1999) y Decano de la GSII en la Universidad de Tokio (2012-2015). Sus áreas de investigación abarcan la informática social, la informática médica y la economía aplicada. Desde 2018, ha representado a Japón en el grupo de expertos en inteligencia artificial de la OCDE (AIGO). Además, ha sido invitado a diversos comités y consejos gubernamentales para asesorar en el desarrollo de políticas públicas. Desde 2016, preside la "Conferencia hacia una Sociedad en Red con IA", organizada por el Ministerio de Asuntos Internos y Comunicaciones de Japón.

> TSURU, KIYOSHI

Abogado líder en el ámbito de la propiedad intelectual, con una especialización en innovación y ciberespacio. Es fundador y director de TMI Law, donde asesora a empresas y organizaciones en la protección y gestión de activos intangibles en el entorno digital. Su experiencia abarca la resolución de disputas sobre nombres

de dominio, la protección de derechos de autor en entornos digitales, software, patentes, indicaciones geográficas, transacciones electrónicas relacionadas con derechos de propiedad intelectual y arbitraje. Además, se desempeña como Country Manager de BSA | The Software Alliance, representando a la organización en México y promoviendo políticas y estrategias para la protección del software y los derechos de propiedad intelectual. También es representante en México de la Entertainment Software Association (ESA), donde actúa como enlace con la industria del entretenimiento digital, asegurando el cumplimiento de regulaciones y promoviendo iniciativas de protección de propiedad intelectual en el sector. Su formación académica incluye un LLM en Derecho de la Propiedad Intelectual por George Washington University, el Internet Law Program de Harvard Law School en el Berkman Center for Internet & Society, el CIDE – Yale US-Mexico Leadership Program, una especialización en Obligaciones y Contratos por la Escuela Libre de Derecho (ELD) y una Licenciatura en Derecho por la Universidad Iberoamericana. Sus áreas de especialización incluyen la propiedad intelectual y el derecho digital, la resolución de disputas sobre nombres de dominio, la protección de derechos de autor en entornos digitales, la protección de software y patentes, las indicaciones geográficas y marcas comerciales, las transacciones electrónicas y contratos digitales, así como el arbitraje y mediación en propiedad intelectual.

> YADAV, ARTI

Profesora Asistente en el Departamento de Comercio y Estudios de Gestión en Ramanujan College, Universidad de Delhi. Obtuvo su Postdoctorado en el Consejo Indio de Investigación en Ciencias Sociales (2019-2021), su Doctorado en Comercio en la Universidad Musulmana de Aligarh (2014-2017), una Maestría en Comercio en la Universidad de Delhi (2013) y su Licenciatura en Comercio en Daulat Ram College, Universidad de Delhi (2011). Sus áreas de docencia incluyen Negocios Internacionales, Economía y Derecho Corporativo, mientras que sus líneas de investigación se centran en comercio internacional, inversión extranjera directa, variables macroeconómicas, crecimiento y desarrollo económico.

> YALOVETZKY, AMIKAM

Gerente Senior de Comunicaciones de Marketing Corporativo en MediaTek. Ingeniero en Electrónica y Comunicaciones de la Escuela de Ingeniería de Jerusalén en Israel. capitán retirado de la Fuerza Aérea de Israel y ha recibido dos premios como "Autor más vendido internacional" de Amazon. Es el Responsable del desarrollo comercial de los mercados de América Latina y Estados Unidos. Con sede en Boston, MA, y con más de 25 años de experiencia en el sector tecnológico.

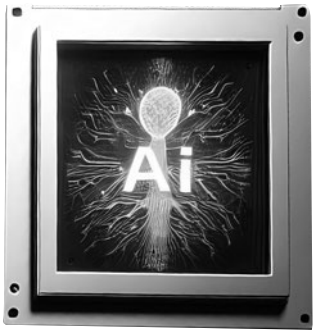
> ZHOU, HUI

Subdirector, profesor asociado. Departamento de Derecho Cibernético e Informático, Instituto de Derecho de la Academia China de Ciencias Sociales (CASS). Su labor académica se centra en la intersección del derecho y las tecnologías de la información, contribuyendo al desarrollo de políticas y regulaciones en el ámbito cibernético en China.

> ZREIK, MOHAMAD

Investigador postdoctoral, Escuela de Estudios Internacionales, Universidad Sun Yat-sen. Doctor en Relaciones Internacionales por la Universidad Normal Central de China (2021). Sus áreas de investigación incluyen la política exterior de China, la Iniciativa de la Franja y la Ruta, estudios sobre Medio Oriente, relaciones China-Árabe, asuntos de Asia Oriental, geopolítica de Eurasia y economía política. Es revisor de varias revistas académicas y miembro del Consejo de Administración del Centro de Estudios Asiáticos y Chinos. Se desempeñó como líder de la comunidad libanesa en Wuhan (octubre 2020 – octubre 2021) y fue miembro de la Federación Internacional de Periodistas y Escritores Árabes, Amigos de China. Sus investigaciones y artículos han sido publicados en revistas de alto impacto y medios internacionales reconocidos. Sus escritos han sido traducidos a múltiples idiomas, incluyendo francés, árabe, español, alemán, albanés, ruso, bosnio y búlgaro.

INTELIGENCIA



ARTIFICIAL

Hacia una nueva era
en la historia de la humanidad

La inteligencia artificial deviene estelar y omnipresente. Los contornos de su geografía no presentan límites. Se eleva desde su plataforma científica a gran velocidad orbitando en paralelos que amenazan alejarse del entendimiento humano.

Los grandes avances y beneficios que la inteligencia artificial puede representar para la humanidad están siendo detallados ampliamente y esta no puede renunciar a ellos de cara a un mundo mejor.

Sin embargo, desde su lanzamiento también empiezan a surgir diferentes voces que alertan de los peligros y las amenazas que un entendimiento equivocado, mal manejo, o una apropiación deshumanizada de la herramienta, representarían no solo para el desarrollo sustentable y equitativo de la sociedad actual, si no para lo más preciado que tiene la humanidad que es su libertad.

El futuro que viene de la mano de la inteligencia artificial no es un debate que pueda resumirse entre tecnófilos y lúditas; seguramente es mucho más que eso.

Debería de ser, entre otras cosas, la posibilidad de resolver con acierto el uso de la mejor herramienta lograda por el ser humano, que es la inteligencia artificial, para que esta a su vez lo acompañe en la construcción de un mundo más justo y sustentable a lo largo de su camino en el siglo XXI.

ISBN: 978-607-99893-1-6

9 786079 989316