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

La voz del INADI Núm. 22

Gobernanza de la Inteligencia Artificial: Hacia una Nueva Era de Responsabilidad y Bien Común

José Ramón López-Portillo Romano

| junio, 2025

Leer

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.

---

* Este ensayo formó parte del libro "Inteligencia artificial. Hacia una nueva era en la historia de la humanidad ", que fue publicado en 2025.

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 apremian-

tes 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. Organizaciones 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.

INADI

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.

JUNIO 2025