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

La voz del INADI Núm. 23

Pasado, presente y futuro de la Inteligencia Artificial

Carlos A. Coello Coello | julio, 2025

Leer

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.

---

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

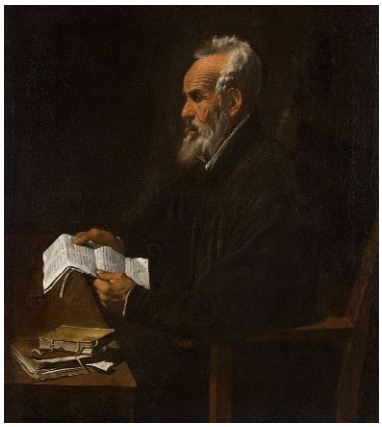
II. Precursores de la Inteligencia Artificial

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.

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.

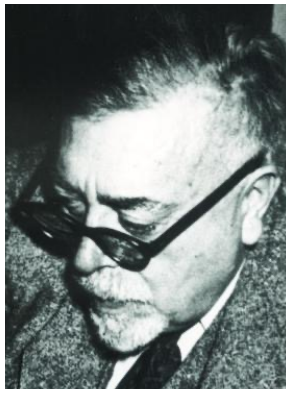
Sin embargo, históricamente, suelen considerarse a tres personajes como los precursores principales de la IA: (1) Norbert Wiener, (2) Claude Shannon y (3) Alan Turing. A continuación, se describirán brevemente las contribuciones principales (relacionadas con la Inteligencia Artificial) realizadas por cada uno de ellos.

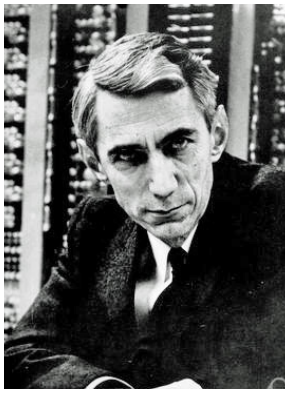
FIGURA 1: Ramón Llull.

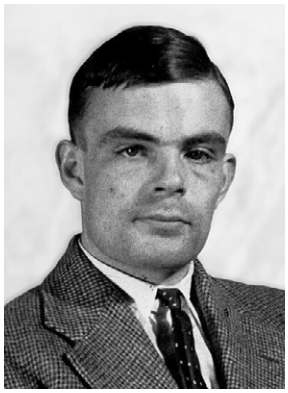


FUENTE: Wikipedia (imagen en el dominio público)

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







FUENTE: Wikipedia (imágenes en el dominio público).

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.

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.

FIGURA 3: Marvin Minsky.

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

III. El nacimiento de una disciplina

FIGURA 4: John McCarthy.

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

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.

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.

FIGURA 5: Frank Rosenblatt.

FUENTE: Cornell University division of rare and manuscript collections.

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

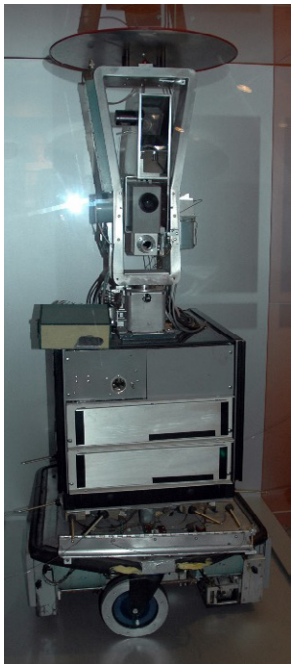
demostró públicamente el 23 de junio de 1960. Rosenblatt describió los detalles de su perceptrón en un artículo de 1958 [Rosenblatt, 1958]. Su modelo consistía de una capa de entrada, una capa oculta (que no podía aprender) con pesos aleatorios y una capa de salida con conexiones que permitían aprender. 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 6: John Alan Robinson.

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

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 7: El robot Shakey.



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.

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*).

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.

FIGURA 9: Shunichi Amari.

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

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)

FIGURA 10: Edward Feigenbaum.

FUENTE: Wikipedia (disponible en el dominio público).

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.

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.

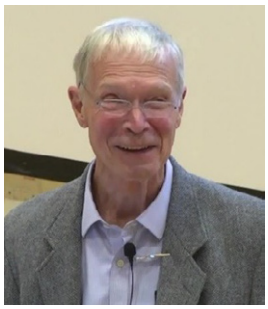
Los sistemas expertos fueron la primera herramienta de IA que se aplicó de manera exitosa en problemas del mundo real. Feigenbaum desarrolló posteriormente sistemas expertos para otras áreas, incluyendo medicina (MYCIN) y genética molecular (MOLGEN).

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.

FIGURA 11: John J. Hopfield.



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

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.

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.

FIGURA 12: David E. Rumelhart.

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

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."

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)

FIGURA 13: Una *LISP Machine*.

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

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 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*.

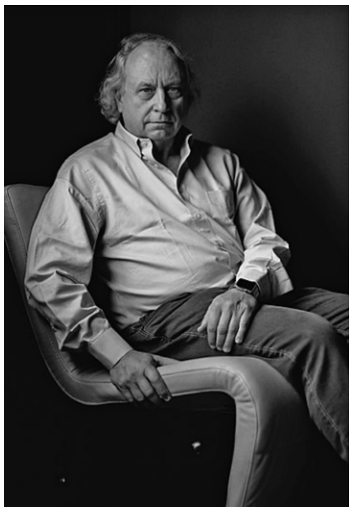
El desempeño de las estaciones de trabajo de *Sun Microsystems* comenzó a poner en aprietos a los desarrolladores de las *LISP Machines*. Pronto, otras empresas como Apple e IBM comenzaron a ofrecer también arquitecturas más simples y populares para ejecutar programas en LISP. Para 1987, estas computadoras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

FIGURA 14: Rodney Brooks.



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

FIGURA 15: Yann LeCun.

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

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 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.

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.

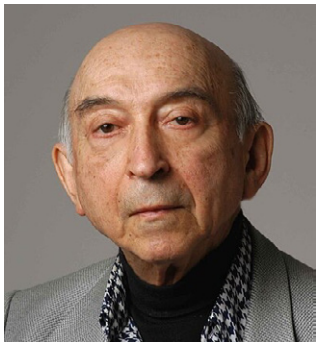
Pero tal vez uno de los eventos más relevantes de este período haya sido la incorporación de múltiples disciplinas (p.ej., matemáticas, economía, investigación de operaciones e ingeniería eléctrica) a la IA.

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

FIGURA 19: Stanley.

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

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.

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

FIGURA 20: Yoshua Bengio.

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

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 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.

INADI

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

JULIO 2025