ENERO 2025
Vol. II, Núm. 1, pp. 093–116
ISSN: 3020-4925

Inteligencia artificial: desde Alan Turing hasta el ChatGPT

Manuel de León

En este artículo, queremos presentar una aproximación histórica a uno de los paradigmas actuales que están influyendo de una manera espectacular en nuestras vidas: la llamada Inteligencia Artificial. Aunque nos pueda parecer algo nuevo, la humanidad ha desarrollado a lo largo de la historia diversas formas de IA que ahora culminan en la actual IA. Se incluyen también algunas nociones técnicas de la IA a fin de acercarla más al lector.

Palabras clave: Inteligencia Artificial, Redes Neuronales, ChatGPT, robots, autómatas.

Artificial Intelligence: from Turing to ChatGPT

In this article, we want to present a historical approach to one of the current paradigms that are dramatically influencing our lives: the so-called Artificial Intelligence. Although it may seem new to us, mankind has developed throughout history various forms of AI that now culminate in the current AI. Some technical notions of AI are also included in order to bring it closer to the reader.

Keywords: Artificial Intelligence, Neural Networks, ChatGPT, robots, automata.

MSC2020: 68T01.

> *«Hay longitudes de onda que la gente no puede ver, hay sonidos que la gente no puede oír y tal vez los ordenadores tengan pensamientos que la gente no pueda pensar. . . »*
>
> Richard Wesley Hamming

Introducción

La Inteligencia Artificial (IA en sus siglas) ha surgido como un fenómeno que pareciera salido de la nada, cuando es un instrumento que venimos usando desde hace décadas. Pero lo que llamamos «IA» ha sufrido una evolución a lo largo del tiempo desde que Alan M. Turing (Londres, 1912 – Wilmslow, 1954) publicara su (entonces polémico) artículo: *Computing Machinery and Intelligence* (Turing, 2012). Turing comienza su obra con estas palabras:

> *«Propongo considerar la siguiente pregunta: '¿Pueden pensar las máquinas?'. Se debiera comenzar definiendo el significado de los términos 'máquina' y 'pensar'. Estas definiciones deberían ser elaboradas de manera tal que reflejen lo mejor posible el uso normal de estas palabras, pero una actitud así es peligrosa. Si el significado de las palabras 'máquina' y 'pensar' proviene del escrutinio de cómo son usadas comúnmente, se hace difícil escapar de la conclusión de que el significado y respuesta a la pregunta '¿pueden las máquinas pensar?' debiera ser buscado en una encuesta estadística, tal como la encuesta Gallup. Pero eso es absurdo. En vez de intentar una definición así, propondré reemplazar esa pregunta por otra, la cual se encuentra estrechamente relacionada y que se puede expresar en palabras relativamente poco ambiguas.»*

Esta reflexión es lo que llevó a Turing a introducir la prueba que hoy se llama *Test de Turing*.

Estas cuestiones de Turing se nos presentan de enorme actualidad, y nos deberían llevar a pensar con detenimiento si lo que ahora estamos llamando IA es en realidad una forma de pensamiento artificial comparable a lo que es capaz de hacer nuestro cerebro (procesos que en gran medida desconocemos) (De León & Timón, 2014).

El propósito de este artículo es mostrar cómo, a lo largo de la historia, la humanidad ha ido creando instrumentos que en cierta manera se pueden considerar antecedentes de lo que se podría llamar inteligencia artificial, y cómo ésta es en definitiva un deseo y un objetivo que hemos perseguido por milenios. En el artículo, se reflexiona también sobre los problemas energéticos que plantea la IA así como sobre los éticos (Degli-Esposti, 2023) y, en particular, sobre su uso en la educación.

Máquinas «inteligentes» antes de Turing

1. Autómatas

Iniciamos nuestro camino hablando de algunos aspectos de la Inteligencia artificial que a veces no identificamos como tales, pero que están en la base de la misma. Hablamos de los autómatas.

Según el Diccionario de la Real Academia Española, autómata es una palabra con diferentes acepciones: [1.] Instrumento o aparato que encierra dentro de sí el mecanismo que le imprime determinados movimientos. [2.] Máquina que imita la figura y los movimientos de un ser animado. [3.] Persona que actúa sin reflexión.

De hecho, la palabra autómata es la latinización del griego antiguo *automaton*, que significa «actuar por voluntad propia».

La figura de los autómatas fue popular en la mitología de la Grecia clásica; a título de ejemplo, Talos era un hombre de bronce que protegía Creta de piratas e invasores. En la Argonáutica de Apolonio de Rodas, se describe a Talos hecho de bronce e invulnerable, con la excepción de una vena en el tobillo que sólo estaba protegida por una fina capa de piel (lo que luego le pasó a Aquiles, por cierto). Talos aparece escenificado en una de las innumerables películas sobre los griegos, en este caso en *Jasón y los argonautas* (véase la figura 1), dirigida por Don Chaffey en 1963.

Pero los griegos no sólo se conformaron con la mitología, sino que también idearon autómatas con sus conocimientos científicos. Uno de los científicos más notables fue Herón de Alejandría. De hecho, es autor de un libro, *Los autómatas*, en el que describe máquinas que permiten realizar en contextos teatrales acciones por medios mecánicos o neumáticos (por ejemplo, apertura o cierre automáticos de puertas de templos, estatuas que vierten vino y leche, etc.)

Figura 1. Talos; Cortesía de Columbia Pictures©.

En la antigua China, hay una descripción extraordinaria en el *Lie Zi* (*ca.* 350 a.C.) en la que un artesano, Yan Shi, le presenta al rey una figura de tamaño natural con forma humana. Según el citado texto: «*El rey contempló la figura con asombro. Caminaba con pasos rápidos, moviendo la cabeza arriba y abajo, de modo que cualquiera la habría tomado por un ser humano vivo. El artífice le tocó la barbilla y empezó a cantar, perfectamente afinada*». El problema surge cuando el autómata se insinúa a las damas de la corte y deben destrozarlo para que el rey vea que está hecho de cuero, madera, cola y laca.

En la edad dorada del califato de Bagdad, se construyeron muchos autómatas: pájaros de metal que cantaban automáticamente batiendo las alas, flautistas, etc. Quizás, el inventor más notable fue Al Jazarí (Cizre -Turquía-, 1136 – ? -Turquía-, 1206), conocido sobre todo por haber escrito *El libro del conocimiento de los ingenios mecánicos* en 1206, donde describía 50 dispositivos mecánicos junto con instrucciones sobre cómo construirlos (véase la figura 2). Uno de sus inventos más famosos es el reloj elefante. Se le ha descrito como el «padre de la robótica» y de la ingeniería moderna. Su tarea era crear máquinas similares a los humanos con fines prácticos, para nuestra comodidad. Así, diseñó varios tipos de sirvientes.

En la Italia del Renacimiento, se dio un resurgimiento del interés por los autómatas, que continuó en siglos posteriores. Un ejemplo que pervive en nuestros días es el reloj astronómico de Praga construido

en 1410 y al que se añadieron luego las figuras que siguen deleitando hoy en día a millares de turistas.

Figura 2. Un *sirviente mecánico* de Al Jazarí.

Cuando René Descartes (La Haye, 1596 – Estocolmo, 1650) indica que los cuerpos de los animales no son más que máquinas complejas, se produce un nuevo estímulo para la creación de autómatas, ya que los huesos, los músculos y los órganos podrían sustituirse por engranajes, pistones y levas. Dejaremos para otra sección el caso de los jugadores automáticos de ajedrez, que merece un estudio aparte. Sí señalaremos que, en tiempos más modernos, los autómatas florecieron. Algún ejemplo es esta maravilla del italiano Innocenzo Manzetti (Aosta, 1826 – *ib.*, 1877), un autómata que tocaba la flauta, con forma de hombre, a tamaño natural, sentado en una silla (véase la figura 3). Dentro de la silla, había palancas, bielas y tubos de aire comprimido que hacían que los labios y los dedos del autómata tocaran la flauta siguiendo un programa grabado en un cilindro similar al de los pianos. El autómata podía interpretar 12 arias diferentes. Durante la interpretación, se levantaba de la silla, inclinaba la cabeza y ponía los ojos en blanco.

Si alguien piensa que no hay IA en estos autómatas, nos remitiremos al artista holandés Theo Jansen (Scheveningen, 1948) y sus estructuras automatizadas (las *strandbeesten*, bestias de playa), que pueden caminar con energía eólica o aire comprimido (véase la figura 4). Jansen afirma que su intención es que evolucionen automáticamente y desarrollen inteligencia artificial, con manadas de estas estructuras vagando libremente por la playa. Según el autor:

> *«Hago esqueletos capaces de caminar con el viento. Con el tiempo, estos esqueletos se han vuelto cada vez mejores para sobrevivir a los elementos, como las tormentas y el agua, y con el tiempo quiero poner estos animales en manadas en las playas, para que vivan su propia vida.»*

La verdad es que, cuando las vemos caminando, pensaríamos que hay una inteligencia que las mueve.

Figura 3. El *autómata* de Manzetti.

Figura 4. Un *strandbeesten*.

2. Robots

Los robots son otra fascinante rama de la IA. Aunque la palabra «robot» pueda ser para la RAE sinónimo de autómata, un robot nos causa más inquietud que un autómata, pues lo consideramos capaz de acciones propias.

1. Robots rebeldes

La palabra «robot» fue usada por el escritor checo Karel Čapek (Malé Svatoňovice, 1890 – Praga, 1938) en su obra teatral *R.U.R.* (*Rossum's Universal Robots*), publicada en 1920. En esta obra, los robots se fabrican con una apariencia humana, como si fueran replicantes de *Blade Runner*, aunque en este caso sin emociones. Pero algunos adquieren conciencia e incitan a una revolución contra la humanidad. El origen de la palabra es curioso. Cuenta Čapek que originalmente había querido llamar a las criaturas *laboři* (por labor, trabajo), pero que no le gustó la palabra y pidió consejo a su hermano Josef, que sugirió *roboti* (siervo).

Para muestra de su apariencia humana, veáse este diálogo de la obra:

> *«(…) DOMIN: Sulla, deja que la señorita Glory te eche un vistazo.*
>
> *HELENA: (se levanta y ofrece su mano) Encantada de conocerte. Debe ser muy duro para ti estar aquí, aislada del resto del mundo. No conozco el resto del mundo, Srta. Glory. Por favor, siéntese.*
>
> *HELENA: (se sienta) ¿De dónde eres?*
>
> *SULLA: De aquí, de la fábrica.*
>
> *HELENA: Oh, usted nació aquí.*
>
> *SULLA: Sí, me hicieron aquí.*
>
> *HELENA: (asombrada) ¿Qué?*
>
> *DOMIN: (riendo) Sulla no es una persona, señorita Glory, es un robot.*
>
> *HELENA: Oh, por favor, perdóname…»*

Hoy en día, se fabrican robots en casi todos los países del mundo, y la Robótica es una pujante rama de la ingeniería, pero el temor a la rebelión de los robots está presente en la sociedad a través de numerosas obras de ficción tanto en la literatura (como es el caso de *R.U.R.*) como en el cine. Veamos algunos ejemplos (más adelante, hablaremos de la IA como singularidad y veremos muchos más casos de robots amenazantes).

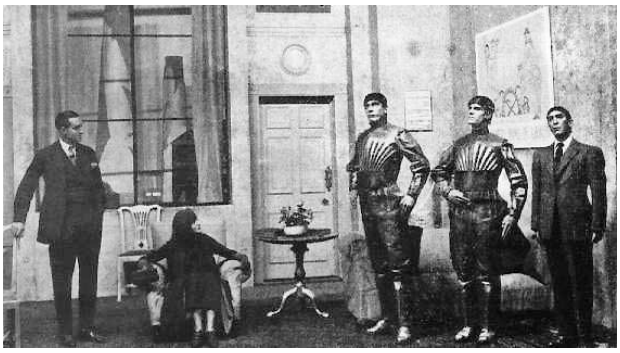


Figura 5. Una escena de la obra en la que aparecen tres robots.

Un primer ejemplo lo vemos en *Saturno 3* (véase la figura 6), una película británica de ciencia ficción de 1980 producida y dirigida por Stanley Donen (Columbia, 1924 – Nueva York, 2019) y protagonizada por Farrah Fawcett (Corpus Christi, 1947 – Santa Mónica, 2009), Kirk Douglas (Nueva York, 1916 – Beverly Hills, 2020) y Harvey Keitel (Nueva York, 1939). El guión fue escrito por Martin Amis (Oxford, 1949 – Lake Worth Beach, 2023) a partir de una historia de John Barry (York, 1933 – Nueva York, 2011).

Figura 6. Cartel de la película *Saturno 3*. Cortesía de ITC Entertainment©.

El argumento es como sigue. En un futuro lejano, una Tierra superpoblada depende de la investigación realizada por científicos en estaciones remotas de todo el Sistema Solar. El capitán Benson (que ha asesinado al capitán encargado de esa misión) parte

hacia una pequeña y remota estación experimental de investigación hidropónica en la tercera luna de Saturno, dirigida por Adam y su ayudante Alex. Benson ensambla un nuevo robot, Héctor, con un cerebro hecho de material cerebral y que está conectado con un enlace neural implantado en su columna vertebral a Benson. El desequilibrio mental de Benson se transfiere a Héctor, así como su interés en la joven Alex, lo que provoca la rebelión del robot y la tragedia final.

Estas palabras de Héctor son esclarecedoras: «*Comprende mi problema, yo soy tú, yo soy Adam, soy el robot, soy todos. La conversación es un arte, yo lo hago a la perfección.*»

Otro robot inquietante es el de la película *Ultimátum a la Tierra*, dirigida por Robert Wise (Winchester, 1914 – Westwood, 2005), basada en una historia de Harry Bates (Pittsburgh, 1900 – Nueva York, 1981) (*El amo ha muerto*; en el original inglés, *Farewell to the Master*, 1940). El argumento gira en torno a un visitante extraterrestre humanoide (Klaatu) que llega a la Tierra, acompañado de un poderoso robot (Gort), para entregar un importante mensaje que afectará a toda la raza humana. Tras una serie de peripecias, consiguen reunir a una serie de científicos a los que Klaatu cuenta que una organización interplanetaria ha creado una fuerza policial de robots invencibles como Gort. «*En cuestiones de agresión, les hemos dado poder absoluto sobre nosotros.*» Klaatu concluye: «*Vuestra elección es sencilla: uniros a nosotros y vivir en paz o seguir vuestro curso actual y enfrentaros a la destrucción. Estaremos esperando vuestra respuesta.*» Klaatu y Gort parten en la nave especial (véase la figura 7).

Figura 7. Klaatu y Gort. Cortesía de 20th Century Studios©.

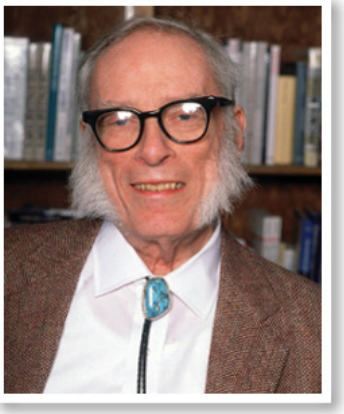
La obra es especialmente interesante para los matemáticos. Klaatu pregunta: «*¿Quién es la persona viva más grande?*», y la sugerencia es un matemático, el profesor Jacob Barnhardt. Cuando visitan su casa, pero éste no está, Klaatu ve que la pizarra de Barnhardt está cubierta de ecuaciones (un intento de resolver el problema de los tres cuerpos). Klaatu corrige sus cálculos.

El relato de Harry Bates es más inquietante. En él, Klaatu es tiroteado y muerto nada más salir de la nave tras decir: «*Soy Klaatu y éste es Gnut*» (Gnut es aquí el nombre del robot). La revelación está en la frase final que pronuncia Gnut: «*No lo entiendes, yo soy el amo.*»

En la próxima sección, hablaremos de los robots buenos y de las leyes de la Robótica que proclamó el gran Isaac Asimov (Petróvichi -Rusia-, 1920 – Nueva York, 1992).

2. *Robots que cumplen las leyes*

Si antes nos hacíamos eco de algunas historias inquietantes de robots, nos fijaremos ahora en cómo podríamos construir robots que fueran útiles y no destructivos, lo que entronca directamente con los desafíos éticos que nos plantea la actual IA.



Isaac Asimov.

En su colección de 1950 *Yo, Robot* (Asimov, 2022), el escritor de origen ruso, nacionalizado estadounidense, incluyó un cuento titulado *Círculo vicioso* (*Runaround* en el original inglés) en el que introdujo las llamadas «Tres Leyes de la Robótica», un conjunto de reglas que debería cumplir un robot. Estas leyes eran:

1. Un robot no puede herir a un ser humano ni, por inacción, permitir que un ser humano sufra daño.
2. Un robot debe obedecer las órdenes que le den los seres humanos, excepto cuando dichas órdenes entren en conflicto con la Primera Ley.
3. Un robot debe proteger su propia existencia siempre que dicha protección no entre en conflicto con la Primera o la Segunda Ley.

En el citado relato (véase la figura 8), el robot Speedy obedece a los dos científicos en su misión, pero tiene un conflicto, ya que no puede decidir si obedecerles (Segunda Ley) o protegerse a sí mismo del peligro (Tercera Ley), así que da vueltas sin tomar la decisión. Finalmente, uno de los científicos se pone en peligro para que se pueda aplicar la Primera Ley y salir así del bucle.

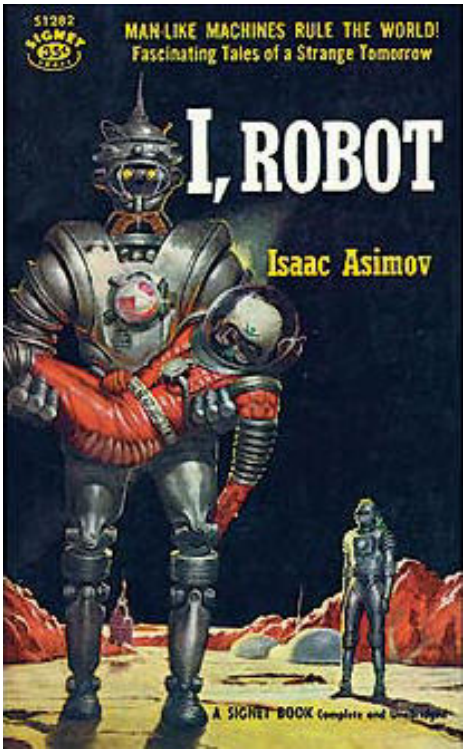


Figura 8. Portada de *I, Robot*.

Estas Tres Leyes fueron sometidas por el autor en sus cuentos de robots a una gran variedad de situaciones que las ponían a prueba.

Asimov presenta dos futuros alternativos y completamente diferentes en las obras *Que te acuerdes de él* y *El hombre bicentenario*, con robots que obvian las Tres Leyes y llegan a considerarse humanos: uno retrata esto de forma positiva con un robot que se une a la sociedad humana, y el otro de forma negativa con robots que suplantan a los humanos.

¿Y qué tienen que ver las Tres Leyes con la Robótica y las IA actuales? Digamos que no las obedecen; los robots más complejos que se fabrican actualmente son incapaces de comprender y aplicar las Tres Leyes. En cuanto a la IA, es un tema que está sobre el tapete. Si la IA aprende (mejorando sus algoritmos) de los datos, lo que está es aprendiendo de nosotros y reproduciendo nuestros sesgos. Es significativa esta opinión del escritor canadiense de ciencia-ficción Robert J. Sawyer (Ottawa, 1960):

> *«El desarrollo de la IA es un negocio, y a las empresas no les interesan las garantías fundamentales, especialmente las filosóficas. (Algunos ejemplos rápidos: la industria tabaquera, la industria automovilística, la industria nuclear. Ninguna de ellas ha dicho desde el principio que sean necesarias salvaguardias fundamentales, todas se han resistido a las salvaguardias impuestas externamente y ninguna ha aceptado un edicto absoluto contra cualquier daño a los seres humanos).»*

Pero también la ciencia se ha hecho eco de las Tres Leyes; a principios de 2011, el Reino Unido publicó un conjunto revisado de 5 leyes, las 3 primeras de las cuales actualizaban las de Asimov, *Principles of robotics. Regulating robots in the real world* (Boden *et al.*, 2011):

1. Los robots son herramientas multiuso. Los robots no deben diseñarse única o principalmente para matar o dañar a los humanos, salvo en interés de la seguridad nacional.
2. Los humanos, no los robots, son los agentes responsables. Los robots deben diseñarse y utilizarse, en la medida de lo posible, respetando la legislación vigente y los derechos y libertades fundamentales, incluida la intimidad.
3. Los robots son productos. Deben diseñarse mediante procesos que garanticen su seguridad.
4. Los robots son artefactos fabricados. No deben diseñarse de forma engañosa para explotar a usuarios vulnerables, sino que su naturaleza de máquina debe ser transparente.
5. Debe ser siempre posible establecer quién es la persona con responsabilidad legal sobre un robot.

Figura 9. Los robots más famosos de la galaxia. Cortesía de Lucasfilm©.

Para terminar, ésta es la respuesta que el propio Asimov dio en alguna ocasión cuando le preguntaban sobre sus Tres Leyes:

«Siempre que alguien me pregunta si creo que mis Tres Leyes de la Robótica servirán para gobernar el comportamiento de los robots, una vez que sean lo bastante versátiles y flexibles como para poder elegir entre distintos comportamientos, tengo la respuesta preparada.

Mi respuesta es: 'Sí, las Tres Leyes son la única forma en que los seres humanos racionales pueden tratar con robots -o con cualquier otra cosa-.'

Pero cuando digo esto, siempre recuerdo (tristemente) que los seres humanos no siempre son racionales.»

Frases para reflexionar.

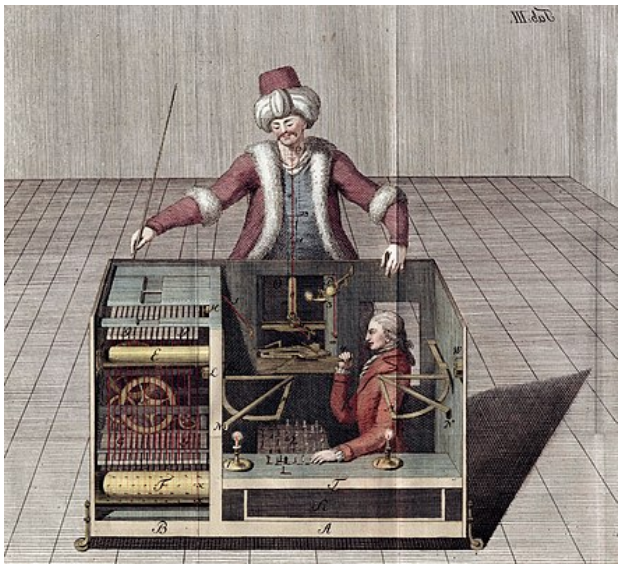


Figura 10. *El Turco* de von Kempelen.

3. Los jugadores de ajedrez

El ajedrez por ordenador se hizo muy popular en las últimas décadas. Uno puede jugar sin que otro jugador humano lo haga contra nosotros, y los programas más sofisticados nos ayudan también a entrenar nuestro juego proponiendo alternativas. Puesto que el ajedrez es sinónimo de inteligencia (ya que el juego ofrece tantas alternativas posibles), no es de extrañar que estos programas de ordenador fueran usados como pruebas de la inteligencia del propio programa.

No nos detendremos aquí en la historia del ajedrez, juego tan querido por los matemáticos. Pero ha habido a lo largo de la historia intentos de crear una máquina para jugar al ajedrez desde el siglo VIII. El intento más famoso se remonta a 1769, con el autómata llamado *El Turco* (véase la figura 10). Fue construido por el escritor e inventor húngaro Wolfgang von Kempelen (Bratislava, 1734 – Viena, 1804). Von Kempelen estudió Derecho y Filosofía en Presburgo y luego en Győr, Viena y Roma, y se interesó también por las matemáticas y la física.

El autómata fue regalado a María Teresa de Austria en 1769. La máquina consistía en un modelo a tamaño natural de una cabeza y un torso humanos, vestidos con túnicas turcas y turbante, sentados detrás de un gran armario sobre el que había un tablero de ajedrez. La máquina parecía capaz de jugar una partida de ajedrez contra un oponente humano, pero en realidad no era más que una elaborada simulación de automatización mecánica: un maestro de ajedrez humano oculto en el interior del armario manejaba al *Turco* desde abajo mediante una serie de palancas. Durante 84 años, varios propietarios la exhibieron en giras como autómata.

Tras la muerte de von Kempelen, su hijo decidió venderlo a Johann Nepomuk Mälzel (Ratisbona -Alemania-, 1772 – La Guaira -Venezuela-, 1838), un músico bávaro interesado en diversas máquinas y dispositivos, quien la exhibió por todo el mundo. En Estados Unidos, Edgar Allan Poe desconfió del autómata y escribió el ensayo *El jugador de ajedrez de Mälzel*, que se publicó en abril de 1836. De todas formas, la literatura sobre este autómata es muy abundante.

Otro famoso intento antes de la era de los ordenadores tiene nombre español, *El Ajedrecista*, de 1912 (véase la figura 11), construido por el ingeniero e inventor Leonardo Torres Quevedo (Santa Cruz de Iguña, 1852 – Madrid, 1936), que jugaba un final de rey y torre contra rey. Este tipo de análisis de finales es muy popular entre los ajedrecistas.

Uno de los algoritmos más famosos para la aplicación de los juegos es el llamado poda alfa-beta, que parece haberse inventado independientemente varias veces. Por ejemplo, el estadounidense Arthur L. Samuel (Emporia, 1901 – Stanford, 1990) tenía una versión temprana para una simulación de damas, y su coterráneo John McCarthy (Boston, 1927 – Stanford, 2011) propuso ideas similares durante la reunión de Dartmouth en 1956 que describiremos más adelante. El nombre de poda viene de que se trata de reducir el número de nodos evaluados en un árbol de juego usando una técnica de minimax muy usada en los procesos de optimización. Arthur Lee Samuel es una figura interesante, ya que creó para el juego de damas ese primer programa de autoaprendizaje con éxito del mundo. A él se le debe además la popularización del término *aprendizaje automático*.

Figura 11. *El Ajedrecista* de Torres Quevedo.

Las aplicaciones informáticas de ajedrez utilizan métodos heurísticos para construir, buscar y evaluar árboles que representan secuencias de jugadas a partir de la posición actual e intentan ejecutar la mejor de dichas secuencias durante la partida. Por lo tanto, la velocidad de cálculo es fundamental. De hecho, el estadounidense Claude Shannon (Petoskey, 1916 – Medford, 2001) dividió los programas en tipo A (fuerza bruta de cálculo) y tipo B (usando inteligencia artificial estratégica).

Las partidas de los programas con los grandes campeones de ajedrez se convirtieron en un espectáculo. En 1997, Deep Blue (véase la figura 12), una computadora de Tipo A, derrotó al Campeón del Mundo Garry Kasparov (Bakú -Azerbaiyán-, 1963). Y en 2005, Hydra, una computadora de ajedrez del Tipo B, derrotó al mejor jugador británico y séptimo mejor clasificado del mundo, Michael Adams (Truro, 1971). Sin embargo, a día de hoy, el ajedrez sigue siendo un juego de una enorme complejidad con millones de posiciones posibles, y no es posible obtener una descripción de una estrategia para el juego perfecto para cualquiera de los bandos.

Remitimos al capítulo final del libro de Benjamín Labatut (2023) para encontrar un debate apasionante sobre la lucha de Kasparov con Deep Blue y la del campeón sudcoreano de Go, Lee Sedol (Sinan, 1983), con AlphaGo Zero.

Figura 12. Deep Blue.

Terminamos esta sección con una lista de nombres que prueba como los grandes pioneros de la IA veían la importancia del ajedrez para su investigación.

En 1948, en su libro *Cybernetics*, el filósofo y matemático estadounidense Norbert Wiener (Columbia, 1894 – Estocolmo, 1964) describe cómo podría desarrollarse un programa de ajedrez utilizando una búsqueda *minimax*.

En 1950, Shannon publica *Programming a Computer for Playing Chess*, uno de los primeros artículos sobre los métodos algorítmicos del ajedrez por ordenador.

En 1951, Turing es el primero en publicar un programa, desarrollado sobre papel, que era capaz de jugar una partida completa de ajedrez (apodado *Turochamp*).

Máquinas que usaban la lógica

Si queremos máquinas que piensen como humanos, y nosotros usamos la lógica, lo natural sería que esta capacidad la comprendiéramos para poder construir esas máquinas inteligentes. Haremos un breve recorrido por algunas figuras históricas que marcaron hitos en el desarrollo de la lógica (que hoy en día es básicamente lógica matemática).

Ramon Llull (Palma de Mallorca, *ca.* 1232 – Mar Mediterráneo, *ca.* 1315) desarrolló varias máquinas

lógicas, entidades mecánicas que podían combinar verdades básicas e innegables mediante simples operaciones lógicas y producir los resultados. Uno de sus principales objetivos era argumentar su oposición a racionalistas como Averroes (Córdoba, 1126 – Marrakech, 1198) y mostrar la verdad desde el punto de vista cristiano. Los sujetos y predicados se organizaban en figuras geométricas consideradas perfectas, como círculos, cuadrados y triángulos. Accionando diales, palancas y manivelas y haciendo girar una rueda, las proposiciones y tesis se desplazaban sobre guías para situarse según el carácter positivo (verdadero) o negativo (falso) que les correspondía, de modo que la máquina podía demostrar por sí misma la verdad o falsedad de un postulado.

En el siglo XVII, Gottfried W. Leibniz (Leipzig, 1646 – Hannover, 1716), Thomas Hobbes (Westport, 1588 – Harwick Hall, 1679) y Descartes exploraron la posibilidad de que todo pensamiento racional pudiera hacerse tan sistemático como el álgebra o la geometría. Hobbes defendía que los seres humanos son puramente físicos y que, por consiguiente, está regido por las leyes del universo. Por ejemplo, escribía en su obra *Leviatán*:

> *«¿Qué es en realidad el corazón si no un resorte; y qué los nervios si no diversas fibras; y qué las articulaciones si no ruedas que dan movimiento a todo el cuerpo?»*

Gottfried Wilhelm Leibniz (1729) y
Thomas Hobbes (*ca.* 1669 – 1670).

Hobbes escribió también:

> *«Porque la razón…no es más que calcular, es decir, sumar y restar.»*

Por su parte, Descartes también fue un mecanicista, basando su trabajo científico en la hipótesis de que los animales son autómatas (bête-máquina). En su obra *Tratado del hombre* dice:

> *«Me gustaría que consideraras que estas funciones (incluida la pasión, la memoria y la imaginación) se derivan de la mera disposición de los órganos de la máquina de forma tan natural como los movimientos de un reloj u otro autómata se derivan de la disposición de sus contrapesos y ruedas.»*

Leibniz creía que gran parte del razonamiento humano podía reducirse a cálculos, y que tales cálculos podían resolver muchas diferencias de opinión:

> *«La única manera de rectificar nuestros razonamientos es hacerlos tan tangibles como los de los matemáticos, de modo que podamos encontrar nuestro error de un vistazo, y cuando haya disputas entre personas, podamos simplemente decir: Calculemos, sin más, a ver quién tiene razón.»*

Leibniz propuso la creación de una characteristica universalis o *característica universal*, construida sobre un alfabeto del pensamiento humano en el que cada concepto fundamental estaría representado por un único carácter real. En palabras suyas:

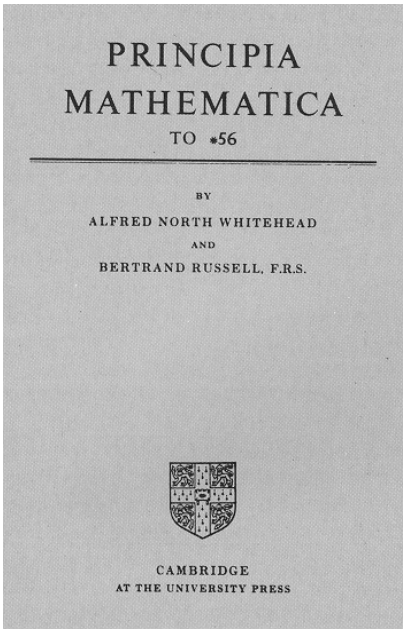
> *«Es evidente que si pudiéramos encontrar caracteres o signos adecuados para expresar todos nuestros pensamientos con la misma claridad y exactitud con que la aritmética expresa los números o la geometría expresa las líneas, podríamos hacer en todas las materias, en la medida en que son susceptibles de razonamiento, todo lo que podemos hacer en aritmética y geometría. Pues todas las investigaciones que dependen del razonamiento se llevarían a cabo por transposición de estos caracteres y por una especie de cálculo.»*

Piano de Jevons.

Uno de los científicos que contribuyó al desarrollo y la aplicación de la lógica fue el inglés William Stanley Jevons (Liverpool, 1835 – Bexhill-on-Sea, 1882), que hizo contribuciones relevantes a la economía aplicando las matemáticas. En el terreno de la lógica, publicó *Pure Logic; or, the Logic of Quality apart from Quantity*, basado en los principios de lógica de Boole, pero liberado de lo que él consideraba el falso ropaje matemático de ese sistema. Con su idea de que «todo lo que es cierto de una cosa es cierto de su semejante», elaboró una máquina lógica, cuyo parecido con un piano vertical determinó su nombre como *piano de Jevons*. Los lectores curiosos pueden consultar el artículo original de Jevons (1870) con los planos para construir su máquina.

La lógica matemática conoce avances decisivos gracias al trabajo de los británicos Bertrand Russell (Trellech, 1872 – Penrhyndeudraeth, 1970) y Alfred North Whitehead (Ramsgate, 1861 – Cambridge -MA, EE.UU-, 1947) en 1913 sobre los fundamentos de las matemáticas en su obra maestra, los *Principia Mathematica*. Los objetivos eran:

1. Analizar en la mayor medida posible las ideas y métodos de la lógica matemática y reducir al mínimo el número de nociones primitivas, axiomas y reglas de inferencia.
2. Expresar con precisión las proposiciones matemáticas en lógica simbólica utilizando la notación más conveniente que permita la expresión precisa.
3. Resolver las paradojas que asolaban la lógica y la teoría de conjuntos a principios del siglo XX, como la paradoja de Russell.



Portada de *Principia Mathematica*.

Otro paso importante es el que da David Hilbert (Königsberg, actual Kaliningrado -Rusia-, 1862 – Gotinga, 1943) con una pregunta esencial: «*¿puede formalizarse todo el razonamiento matemático?*». En efecto, Hilbert propuso un proyecto de investigación en metamatemáticas que se conoció como el programa de Hilbert: formular las matemáticas sobre una base lógica sólida y completa. Creía que, en principio, esto podría lograrse demostrando que todas las matemáticas se deducen de un sistema finito de axiomas correctamente elegido, y que dicho sistema de axiomas es demostrablemente consistente.

Pero nada más pronunciar su célebre frase: «*Wir müssen wissen. Wir werden wissen*» (en español «*Debemos saber, sabremos*»), su edificio se tambaleó con los *Teoremas de incompletitud* del lógico austriaco Kurt Gödel (Brno -Chequia-, 1906 – Nueva Jersey, 1978). Dichos teoremas demuestran que incluso los sistemas axiomáticos elementales, como la aritmética de Peano, son autocontradictorios o contienen proposiciones lógicas imposibles de demostrar o refutar dentro de ese sistema. Por lo tanto, había límites a lo que la lógica matemática podía lograr.

Poco después, Turing introdujo el concepto de *máquina de Turing* en su artículo «On computable numbers, with an application to the Entscheidungsproblem», en 1936, en el que se estudiaba la cuestión planteada por Hilbert sobre si las matemáticas son decidibles, es decir, si hay un método definido que pueda aplicarse a cualquier sentencia matemática y que nos diga si esa sentencia es cierta o no. Así demostró que existían problemas que una máquina no podía resolver.

La revolución de Turing

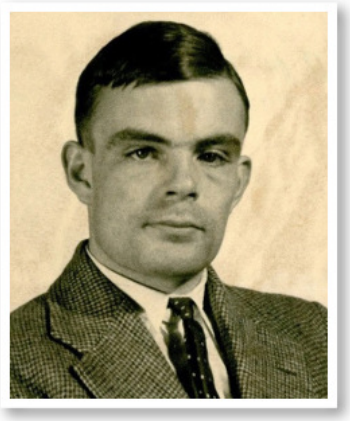
El matemático inglés tuvo una influencia enorme en el desarrollo de las matemáticas, la lógica y las ciencias de la computación. Nuestro mundo no sería igual sin sus enormes contribuciones (Copeland, 2013; De León & Timón, 2014). Desgraciadamente, su papel no fue tan conocido en el mundo de las matemáticas.

Los grandes logros de Turing fueron los siguientes:

- La máquina de computación universal.
- Su trabajo en criptografía.
- Sus aportaciones a la inteligencia artificial.
- La teoría de la morfogénesis.
- Sus aportaciones al análisis numérico.

En 1931, Turing entró en el King's College de la Universidad de Cambridge para estudiar matemáticas.

Aunque no consiguió en primera instancia una beca, repitió el examen al año siguiente para conseguirla. En la universidad, Turing se sintió más libre para explorar su propio camino. Por ejemplo, leyó la *Introducción a la filosofía matemática*, de Russell, o el libro de mecánica cuántica de John von Neumann (Budapest, 1903 – Bethesda -Maryland, EE.UU-, 1957). Fue en 1933 cuando Turing comenzó a interesarse por la lógica matemática.



Alan M. Turing (1936).

Turing se graduó en 1934, y en 1935 asistió a un curso avanzado que dictó Max Newman (Londres, 1897 – Cambridge, 1984) en la Universidad de Cambridge sobre los fundamentos de las matemáticas, basado en gran medida en los trabajos y preguntas de Hilbert. Éstas eran las grandes cuestiones:

1. ¿Son completas las matemáticas? O dicho de otra manera, ¿se puede probar que una proposición es verdadera o falsa sin ningún género de dudas?
2. ¿Son consistentes las matemáticas? ¿No ocurrirá que pueda demostrarse que una proposición es verdadera y falsa a la vez?
3. ¿Son computables? ¿Se puede diseñar un procedimiento mecánico que, partiendo de una proposición, tras un número finito de pasos nos dé la conclusión de si es cierta o falsa?

El seminario concluía con una demostración del *Teorema de incompletitud* de Gödel. Recordemos que Gödel había demostrado que la propia aritmética era incompleta y que su consistencia no podía probarse dentro de su propio marco axiomático, y también que la segunda mitad del siglo XIX y la primera del siglo XX vivieron la gran crisis de los fundamentos de las matemáticas. Newman sembró en Turing vientos que habrían de dar lugar no a tempestades sino a algunos de los logros más brillantes del pensamiento humano. Turing se centró en el tercer tema, el de la computabilidad. Dada una proposición, ¿es posible diseñar un algoritmo que decida si la proposición es cierta o falsa? En algunos casos es fácil encontrar un algoritmo, pero en otros no existe tal algoritmo. Pero es que ni siquiera existía una definición rigurosa de algoritmo, así que Turing comenzó a trabajar en esos temas.

En 1936, Turing publicó otro trabajo notable: «On computable numbers, with an application to the Entscheidungsproblem». El *problema de decisión* (*Entscheidungsproblem*) fue discutido por Hilbert en 1928. En este artículo, introdujo la noción de lo que hoy se denomina *máquina de Turing*. Turing reformuló los resultados de Gödel de 1931, reemplazando el lenguaje formal basado en la aritmética de Gödel por el concepto de máquina de Turing.

Es un constructo abstracto que se mueve de un estado a otro siguiendo un número finito de reglas concretas. La máquina de Turing puede escribir un símbolo en la cinta o borrarlo. Una máquina de Turing sería capaz de realizar cualquier computación matemática si ésta se representa como un algoritmo. Turing probó que no hay solución al *Entscheidungsproblem* demostrando primero que el problema de la parada es indecidible para una máquina de Turing; en general, no es posible decidir algorítmicamente si una máquina de Turing se detendrá.

Alonzo Church (Washington D. C., 1903 – Hudson, 1995) había publicado el artículo «An unsolvable problem in elementary number theory», en *American Journal of Mathematics* en 1936, que probaba algo similar, aunque con una aproximación diferente. Esto provocó algunos problemas para que el artículo de Turing fuera publicado en los *Proceedings of the London Mathematical Society*, y tuvo que hacer algunos cambios y referirse al artículo de Church. Finalmente, se publicó en 1937. Una consecuencia positiva fue que Turing se trasladó en 1936 como estudiante graduado a Princeton, donde trabajó con Church en el Institute for Advanced Studies, regresando a Inglaterra en 1938. En Princeton, realizó su tesis doctoral bajo la dirección de Church, defendiéndola en junio de 1938 con el titulo *Systems of logic based on ordinals*. En su tesis, introdujo el concepto de lógica ordinal y de computación relativa, con conceptos que permitían estudiar problemas no abordables con una máquina de Turing.

En 1937, volvió a pasar sus vacaciones a Inglaterra y conoció al filósofo Ludwig Wittgenstein (Viena, 1889 – Cambridge, 1951), asistiendo a su curso sobre los fundamentos de las matemáticas. Debatieron sobre este tema, pues mientras Turing defendía el

formalismo matemático, Wittgenstein decía que las matemáticas no descubrían verdades absolutas, sino que las inventaban.

En Princeton, Turing había desarrollado la teoría, pero a su vuelta a Cambridge en 1938 comenzó a construir un artefacto con el que pretendía estudiar la hipótesis de Riemann. Pero es entonces cuando el gobierno lo contrata para la escuela de códigos y cifrado y le piden ayuda para romper los códigos alemanes de la máquina Enigma. Así, al estallar la Segunda Guerra Mundial, Turing comienza a trabajar para el ejército en Bletchley Park.

Al final de la guerra, Turing fue invitado por el National Physical Laboratory de Londres para diseñar un ordenador. Redactó un informe en marzo de 1946 proponiendo el *ACE* (*Automatic Computing Engine*), aunque hubo dudas en la viabilidad de este proyecto, y sufrió así retrasos hasta su aprobación definitiva, con lo que Turing no participó ya en la construcción de la máquina piloto. Se supone que muchas de las ideas que von Neumann desarrolló en los Estados Unidos estaban basadas en las de Turing.

En 1948, Newman le ofreció un puesto en la Universidad de Manchester, donde colaboró con Frederic C. Williams (Stockport, 1911 – Mánchester, 1977) y Tom Kilburn (Dewsbury, 1921 – Manchester, 2001) en la construcción de una máquina de computación.

En 1950, Turing publicó un artículo clave para el futuro desarrollo de la Inteligencia Artificial: «Computing machinery and intelligence» (Turing, 2012). En este artículo, Turing propuso lo que hoy se denomina el test de Turing, que se utiliza para conocer si una máquina puede ser inteligente como un ser humano. Esto es la base de los llamados test CAPTCHA, que intentan saber si el usuario es una máquina o un humano.

Post-Turing

En la década de los años 1950, coinciden temporalmente una serie de ideas que son las que darán lugar al nacimiento de la IA:

- La neurociencia había probado a comienzos del siglo XX que el cerebro es una red eléctrica de neuronas que se disparaban en pulsos de todo o nada.
- La cibernética de Wiener describía el control y la estabilidad de las redes eléctricas.
- Shannon había sentado las bases matemáticas de la comunicación. Shannon también introdujo formalmente el término «bit».
- La teoría de la computación de Turing demostró que cualquier forma de computación podía describirse digitalmente.

Redes neuronales artificiales

Por otra parte, en 1943, Walter Pitts (Detroit, 1923 – Cambridge, 1969) y Warren McCulloch (Orange, 1898 – Cambridge, 1969) analizaron redes de neuronas artificiales idealizadas y mostraron cómo podían realizar funciones lógicas sencillas. Fueron los primeros en describir lo que más tarde los investigadores llamarían una *red neuronal*, de las que hablaremos más tarde en extensión.

Uno de los estudiantes inspirados por Pitts y McCulloch fue Marvin Minsky (Nueva York, 1927 – Boston, 2016), que por aquel entonces era un estudiante de posgrado de 24 años. En 1951 Minsky y Dean Edmonds construyeron la primera máquina de red neuronal, el *SNARC* (*Stochastic Neural Analog Reinforcement Calculator*), una simulación por computadora de la forma en que funcionan los cerebros orgánicos. La *SNARC* (véase la figura 13) aprendió de la experiencia y se usó para buscar la salida de un laberinto, como una rata en un experimento de psicología.

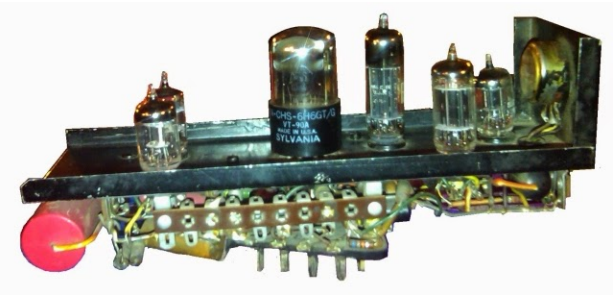
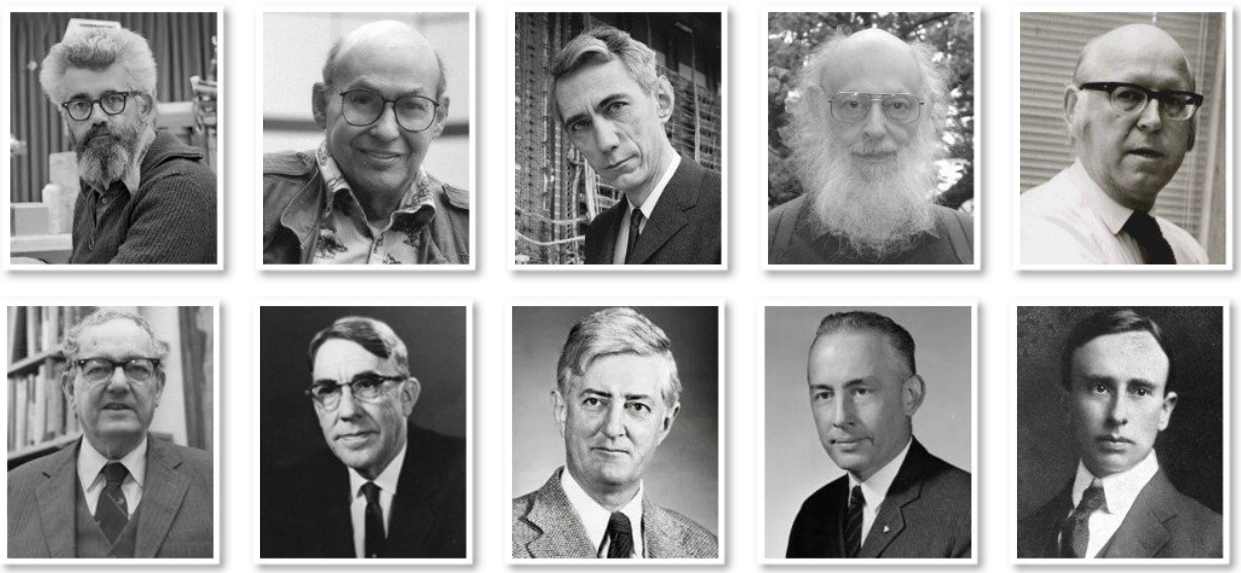


Figura 13. Una de las neuronas del *SNARC*.

Nacimiento de la IA

La Conferencia de Dartmouth sobre Inteligencia Artificial, celebrada en el verano de 1956 durante ocho semanas se considera como el evento en el que se funda la inteligencia artificial. Fueron cuatro investigadores de gran prestigio los organizadores del evento: los anteriormente citados Shannon, Minsky y McCarthy, junto con Nathaniel Rochester (Nueva York, 1919 – Newport, 2001).

Dartmouth College es una universidad privada de las denominadas de investigación, y es una de las nueve universidades coloniales fundadas antes de la Revolución Americana. Dartmouth está considerada una de las universidades más prestigiosas de Estados Unidos. Se eligió esta universidad porque



Conferencia de Darthmouth (1956): Los fundadores de la IA. De izq. a drch. y de arriba a abajo: John McCarthy, Marvin Minsky, Claude Shannon, Ray Solomonoff, Alan Newel, Herbert A. Simon, Arthur Samuel, Oliver Selfridge, Nathaniel Rochester y Trenchard More.

McCarthy, profesor adjunto de Matemáticas en Dartmouth, fue el que tomó la iniciativa. El objetivo era intercambiar ideas con los investigadores más representativos en lo que entonces era la cibernética, teoría de autómatas, teoría de la información, y acuñó un nuevo nombre para el campo: Inteligencia Artificial. Según se comenta en Mitchell (2024), McCarthy tuvo bastantes dificultades en convencerlos a todos para venir y mucha paciencia para que tantos egos juntos pudieran trabajar coordinadamente.

La propuesta original presentada el 31 de agosto de 1955 comenzaba así:

> «*Proponemos que durante el verano de 1956 se lleve a cabo en el Dartmouth College de Hanover, New Hampshire, un estudio sobre la inteligencia artificial, con una duración de 2 meses y 10 participantes. El estudio se basará en la conjetura de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse con tanta precisión que se puede hacer que una máquina lo simule. Se intentará averiguar cómo hacer que las máquinas utilicen el lenguaje, formen abstracciones y conceptos, resuelvan tipos de problemas ahora reservados a los humanos y se mejoren a sí mismas. Creemos que se puede lograr un avance significativo en uno o más de estos problemas si un grupo de científicos cuidadosamente seleccionados trabajan juntos en ello durante un verano.*» (McCarthy *et al.*, 1955)

Durante este congreso, se presentó el programa *Logic Theorist*, escrito por Allen Newell (San Francisco, 1927 –Pittsburgh, 1992), Herbert A. Simon (Milwakee, 1916 – Pittsburgh, 2001) y Cliff Shaw (South California, 1922 – Los Ángeles, 1991). Fue el primer programa diseñado deliberadamente para realizar razonamientos automatizados, y ha sido descrito como el primer programa de inteligencia artificial. Logic Theorist demostró 38 de los 52 primeros teoremas del capítulo dos de *Principia Mathematica* de Whitehead y Russell y encontró pruebas nuevas y más cortas para algunos de ellos. Sin duda, un logro excepcional. Sin embargo, no obtuvo el eco que merecía, tuvo una acogida tibia, probablemente porque, salvo los propios Newell y Simon, nadie se dio cuenta de la importancia a largo plazo de lo que estaban haciendo. Simon confiesa que «probablemente éramos bastante arrogantes al respecto» y añade:

> «*No querían saber nada de nosotros, y nosotros tampoco queríamos saber nada de ellos: ¡teníamos algo que enseñarles!... En cierto modo era irónico, porque ya habíamos hecho el primer ejemplo de lo que buscaban y, en segundo lugar, no le prestaron mucha atención.*»

1. Optimismo inicial

Los programas desarrollados en los años posteriores fueron extraordinarios. Los ordenadores resolvían problemas de álgebra, demostraban teoremas de geometría y aprendían a hablar inglés.

Los investigadores expresaron un intenso optimismo en privado y en la prensa, prediciendo que se construiría una máquina totalmente inteligente en menos de 20 años.

Agencias gubernamentales como la Defense Advanced Research Projects Agency (DARPA, entonces conocida como «ARPA») invirtieron dinero en este campo.

A finales de los 50 y principios de los 60 se crearon laboratorios de Inteligencia Artificial en varias universidades británicas y estadounidenses.

2. Algunas predicciones:

1958, Herbert A. Simon y Allen Newell:

«(…) dentro de diez años un ordenador digital será el campeón mundial de ajedrez»

y

«(…) dentro de diez años un ordenador digital descubrirá y demostrará un nuevo teorema matemático importante.»

1965, Simon:

«(…) las máquinas serán capaces, dentro de veinte años, de hacer cualquier trabajo que pueda hacer un hombre.»

1967, Minsky:

«Dentro de una generación…el problema de crear 'inteligencia artificial' estará sustancialmente resuelto.»

1970, Minsky (en la revista *Life*):

«(…) En un plazo de tres a ocho años tendremos una máquina con la inteligencia general de un ser humano medio.»

Los inviernos de la IA

El término de *invierno de la IA* hace referencia a un periodo de tiempo en el que se produce una drástica reducción de la financiación y del interés por la investigación en este campo, generalmente debido a la decepción por los resultados obtenidos respecto a los esperados.

El término apareció por primera vez en 1984 como tema de un debate público en la reunión anual de la Asociación Americana de Inteligencia Artificial. En esa reunión, dos de los históricos fundadores de la IA, Roger Schank (Nueva York, 1946 – Shelburne, 2023) y Marvin Minsky, advirtieron a la comunidad empresarial de que el entusiasmo por la IA se había descontrolado en los años 80 y de que, sin duda, seguiría la decepción. Describieron una reacción en cadena, similar a un invierno nuclear, que comenzaría con el pesimismo en la comunidad de la IA, seguido del pesimismo en la prensa, seguido de un grave recorte de la financiación, seguido del fin de la investigación seria. Esto fue lo que ocurrió apenas tres años más tarde, cuando la industria de la IA comenzó a derrumbarse.

Históricamente, pueden acotarse dos periodos de invierno.

1. Primer invierno de la IA (1974 – 1980)

Son varios los hitos que marcaron este primer invierno. Por ejemplo, el llamado *Informe Lighthill* que el inglés Sir Michael James Lighthill (París, 1924 – Brecqhou, 1998) presentó al parlamento británico a petición de esta institución y en el que se planteaba el fracaso absoluto de la IA en la consecución de sus grandiosos objetivos. Este informe llevó a un total desinterés en el país por la IA, ya que lo que ésta prometía podría hacerse con otras ciencias.

El Informe Lighthill y el que la misma DARPA desarrolló sugirieron que era improbable que la mayor parte de la investigación en IA produjera algo realmente útil en un futuro previsible. Y en DARPA se había ya instaurado el principio de que sólo se financiarían proyectos dirigidos a resolver problemas concretos y no investigación básica. Recordemos que DARPA es la Agencia de Proyectos de Investigación Avanzada de Defensa para la investigación y desarrollo del Departamento de Defensa de Estados Unidos, y por tanto responsable del desarrollo de tecnologías emergentes para uso militar. DARPA había apostado con fuertes inversiones en el nuevo campo, y su negativa a seguirlo financiando fue decisiva.

2. Segundo invierno de la IA (1987 – 2000)

Un estudio de los informes de principios de la década de 2000 sugiere que la reputación de la IA seguía siendo mala. He aquí algunas informaciones en los medios de comunicación corroborando que, de nuevo, las expectativas no se habían hecho realidad:

Alex Castro, citado en *The Economist*, 7 de junio de 2007:

«(Los inversores) se sentían desalentados por el término reconocimiento de voz que, al igual que inteligencia artificial, se asocia a sistemas que con demasiada frecuencia no han cumplido sus promesas.»

Patty Tascarella en el *Pittsburgh Business Times*, 2006:

«Algunos creen que la palabra robótica conlleva un estigma que perjudica las posibilidades de financiación de una empresa.»

John Markoff en el *New York Times*, 2005:

«En su momento más bajo, algunos informáticos e ingenieros de software evitaron el término inteligencia artificial por miedo a ser vistos como soñadores con ojos salvajes.»

Redes neuronales

Para esta sección, se han consultado varias referencias, como Ruiz-Sarrias (2024) o Ríos-Insua y Gómez-Ullate (2019), además de algunos artículos publicados en la red.

1. Nociones generales sobre redes neuronales

Las redes neuronales reciben este nombre porque están inspiradas en la estructura de las redes neuronales de nuestro cerebro. El descubrimiento de Santiago Ramón y Cajal (Petilla de Aragón, 1852 – Madrid, 1934), Premio Nobel de Medicina o Fisiología en 1906, de cómo las unidades fundamentales del cerebro eran las neuronas, y la intrincada relación entre ellas (vía transmisores químicos y descargas eléctricas) parecía abrir el camino en el intento de reproducir la formación del pensamiento humano.

Así, en el aprendizaje automático (que es en definitiva lo que estamos llamado ahora IA), una red neuronal se compone de unidades que modelan las neuronas del cerebro. Estas neuronas artificiales están conectadas entre sí por bordes que imitan las sinapsis del cerebro, de modo que cada neurona artificial recibe señales de las neuronas conectadas, las procesa y envía una señal a otras neuronas conectadas. Las señales que en el caso de las neuronas biológicas son transmisores químicos o descargas eléctricas, se simulan ahora mediante dos funciones esenciales, la función de propagación y la función de activación.

Hay unos valores de entrada que se especifican en vectores, así como sus pesos correspondientes y un valor de sesgo, de modo que la función de propagación f_{pr} depende de todos estos valores

$$z = f_{\text{pr}}(x, w, b) = \sum_{i=1}^{n} x_i w_i + b.$$

La intensidad de la señal en cada conexión viene determinada por esos pesos, que se van ajustando durante el proceso de aprendizaje.

La función de activación f_{act} depende de la salida de la función de propagación. Por ejemplo, podría ser calcular el máximo entre 0 y z, de manera que si el valor de salida es negativo da 0 y si es positivo, deja z sin variar. En resumen, la neurona recibe la entrada y los pesos, añade el sesgo, y de allí sale $f_{\text{act}}(z)$.

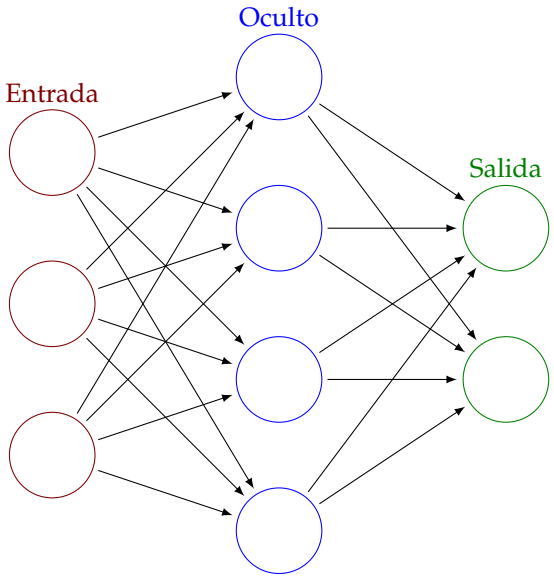


Figura 14. Un ejemplo de red neuronal

Esto es lo que sería la actuación en una única neurona artificial, pero las redes neuronales se componen de capas. Las distintas capas pueden realizar diferentes transformaciones en sus entradas. Las señales viajan desde la primera capa (la capa de entrada) hasta la última (la capa de salida), posiblemente pasando por múltiples capas intermedias (capas ocultas). Una red suele denominarse red neuronal profunda si tiene al menos dos capas ocultas. Cada capa está formada por varias neuronas, de manera que todas las neuronas de una capa están conectadas con todas las de la siguiente capa. Digamos que la capa de entrada se conecta a las ocultas que son intermedias, y éstas últimas a la capa de salida. Así, el potencial de cada neurona se amplifica enormemente.

2. Entrenamiento de una red neuronal

El entrenamiento de la red neuronal es el proceso más relevante si queremos construir redes que sean eficientes y que nos den las respuestas adecuadas. Tenemos que distinguir dos tipos de salidas: un valor numérico o una categoría. Según cada caso, se usará una función de activación u otra. La forma de entrenar la red es proporcionarle datos de entrada y salida fiables para que se puedan hacer las correcciones pre-

cisas.

Para ello, necesitamos una nueva función, que viene de la teoría de optimización y que se denomina la función de coste. Una de estas funciones es, por ejemplo, la de error cuadrático medio. Esta técnica se conoce desde hace más de dos siglos como método de mínimos cuadrados o regresión lineal, y fue utilizada por Carl F. Gauss (Brunswick, 1777 – Gotinga, 1855) para encontrar un buen ajuste lineal aproximado a un conjunto de puntos para predecir el movimiento planetario (Gauss lo usó para predecir el movimiento de Ceres, el primer asteroide conocido, catalogado ahora como planeta enano).

Las redes neuronales que utilizan una función de coste de error cuadrático medio pueden emplear métodos estadísticos formales para determinar la confianza del modelo entrenado, estimando la varianza. Este valor puede utilizarse para calcular el intervalo de confianza de la salida de la red, suponiendo una distribución normal.

La función de coste nos dará el error cometido entre lo esperado y el valor obtenido. Y aquí vuelven a entrar en juego los pesos y los sesgos. Dos de los algoritmos usados popularmente son el método del descenso y el del gradiente, bien conocidos en los ámbitos de la optimización y el análisis numérico.

3. Redes neuronales especiales

La arquitectura general de una red neuronal descrita previamente puede modificarse para tratar tareas más específicas. Es el caso de las redes neuronales convolucionales y las redes neuronales recurrentes.

La arquitectura de redes neuronales convolucionales fue introducida por Kunihiko Fukushima (Taiwan, 1936) en 1979, y es clave para tratar la visión por ordenador. En efecto, funcionan muy bien cuando se dispone de datos organizados espacialmente, como pasa con imágenes. Si una imagen está formada por píxeles, la operacion de convolución permite analizar cada píxel y su relación con los píxeles próximos a éste. La operación se efectúa con las llamadas matrices de convolución, que, simplificando el proceso, actúan como sigue: cada imagen se representa por una matriz I, existe otra llamada *el núcleo K* (que da precisamente la información del contorno), y la operación de convolución entre estas dos matrices da el filtrado de la imagen. Es un procedimiento bien conocido por los que trabajan el procesado de imágenes. Lo que hacen estas redes neuronales es imitar el funcionamiento del córtex visual.

Las redes neuronales recurrentes tratan, por el contrario, de imitar los procesos de la lectura, en los que no sólo leemos las palabras en sí, sino que tenemos un contexto en nuestra memoria que hace que entendamos lo que estamos viendo. Para ello, se introduce un parámetro que funciona como una memoria de la red.

El aprendizaje profundo (deep machine learning) utiliza múltiples capas de neuronas entre las entradas y las salidas de la red. Éstas pueden extraer progresivamente características de nivel superior a partir de la entrada bruta. Por ejemplo, en el procesamiento de imágenes, las capas inferiores pueden identificar bordes, mientras que las capas superiores pueden identificar los conceptos relevantes para un ser humano, como dígitos, letras o caras.

Digamos que el éxito arrollador del aprendizaje profundo no se debe a un nuevo descubrimiento o avance teórico, ya que las redes neuronales proceden de la década de 1950, sino a dos factores: el increíble aumento de la potencia de los ordenadores y la disponibilidad de enormes cantidades de datos de entrenamiento.

Hasta aquí, hemos descrito lo esencial de las redes neuronales. La cantidad de trabajo desarrollado y que está desarrollándose es espectacular y seguramente veremos logros que parecían hasta hace poco impensables. Pero como todo, la IA tienen sus peligros, como veremos a continuación.

La espectacular primavera actual de la IA

En las primeras décadas del siglo XXI, el acceso a grandes cantidades de datos, ordenadores más baratos y rápidos y técnicas avanzadas de aprendizaje automático se aplicaron con éxito a muchos problemas en toda la economía.

Un punto de inflexión fue el éxito del aprendizaje profundo en torno a 2012, que mejoró el rendimiento del aprendizaje automático en muchas tareas, como el procesamiento de imágenes y vídeos, el análisis de textos y el reconocimiento del habla.

Esa interacción humano-máquina es uno de los grandes éxitos de la IA: nuestros ordenadores nos hablan y podemos establecer un diálogo con ellos. Estos avances tienen sus inicios en programas pioneros como *ELIZA*, uno de los primeros programas informáticos de procesamiento del lenguaje natural desarrollado entre 1964 y 1967 en el MIT por Joseph

Weizenbaum (Berlín, 1923 – Ludwigsfelde, 2008). ELIZA simulaba conversaciones utilizando una metodología de sustitución y concordancia de patrones que daba a los usuarios una ilusión de comprensión por parte del programa, pero no tenía una representación que pudiera considerarse realmente comprensiva de lo que decía cualquiera de las partes. ELIZA fue uno de los primeros chatbots capaces de intentar pasar el test de Turing.

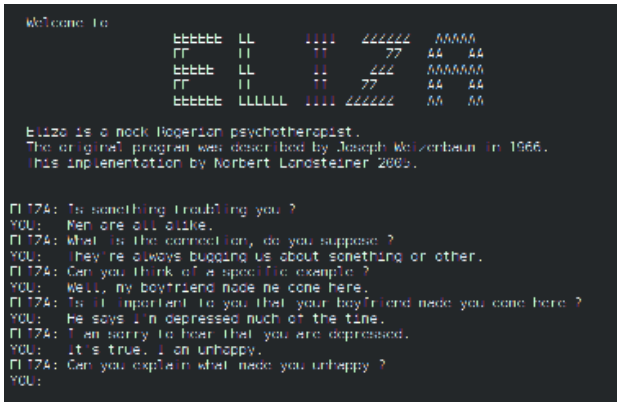


Figura 15. Una conversación con ELIZA.

En 2002, la preocupación era que la IA había abandonado en gran medida su objetivo original de producir máquinas versátiles y totalmente inteligentes, y abogaron por una investigación más directa de la inteligencia general artificial.

Durante el mismo periodo, los nuevos conocimientos hicieron temer que la IA fuera una amenaza existencial.

1. Aprendizaje profundo

En 2012, *AlexNet*, un modelo de aprendizaje profundo, desarrollado por el canadiense de origen ucraniano Alex Krizhevsky, ganó el *ImageNet Large Scale Visual Recognition Challenge*.

El aprendizaje profundo utiliza un perceptrón multicapa. Aunque esta arquitectura se conoce desde los años 60, para que funcione se necesita un hardware potente y grandes cantidades de datos de entrenamiento. Geoffrey Hinton (Londres, 1947) recuerda que, en los años 90, el problema era que *«nuestros conjuntos de datos etiquetados eran miles de veces demasiado pequeños. Y nuestros ordenadores eran millones de veces demasiado lentos»*.

Hinton es conocido por sus trabajos sobre redes neuronales artificiales, que le valieron el título de «Padrino de la IA». Antes de que esto estuviera disponible, la mejora del rendimiento de los sistemas de procesamiento de imágenes requería características *ad hoc* creadas a mano que eran difíciles de implementar.

A principios de la década de 2020, la IA ha alcanzado los niveles más altos de interés y financiación de su historia según todos los indicadores posibles: publicaciones, solicitudes de patentes, inversión total (50 000 millones de dólares en 2022), ofertas de empleo (800 000 ofertas de empleo en EE.UU. en 2022).

Los éxitos de la actual *primavera de la IA* o *boom de la IA* son: los avances en la traducción de idiomas, el reconocimiento de imágenes, los juegos virtuales como *AlphaZero* (campeón de ajedrez) y *AlphaGo* (campeón de go) y *Watson* (campeón de Jeopardy).

El lanzamiento en 2022 del chatbot de IA *ChatGPT* de *OpenAI*, que en enero de 2023 contaba con más de 100 millones de usuarios, ha revitalizado el debate sobre la inteligencia artificial y sus efectos en el mundo.

¿Sabremos usar la IA adecuadamente o será una amenaza?: la singularidad

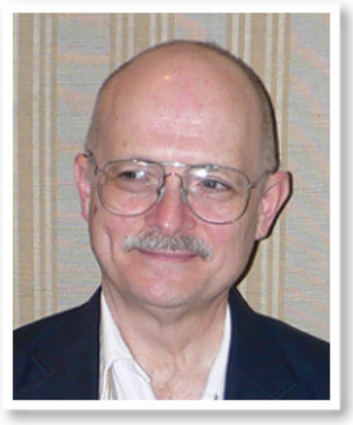
Uno de los peligros que algunos señalan para el futuro próximo de la IA es la llamada singularidad, es decir, el momento en que aparezca una inteligencia artificial muy superior a la de la mente humana. Si la investigación en inteligencia general artificial produjera un software suficientemente inteligente, podría ser capaz de reprogramarse y mejorarse a sí mismo. El software mejorado sería aún mejor para retroalimentarse. Este tema es popular en los medios de comunicación y también entre la población, ya que la literatura y el cine lo ha difundido en muchas obras.

Esa singularidad tecnológica es ese hipotético punto futuro en el tiempo en el que el crecimiento tecnológico se vuelve incontrolable e irreversible. Según Irving J. Good (Londres, 1916 – Radford -EE.UU-, 2009), lo que él llamó el *modelo de explosión de inteligencia*, un agente inteligente mejorable podría entrar en un bucle de retroalimentación positiva de ciclos de autosuperación, en el que cada generación sucesiva y más inteligente aparecería más y más rápidamente, provocando un rápido aumento («explosión») de la inteligencia que, en última instancia, daría lugar a una poderosa superinteligencia, cualitativamente muy superior a toda la inteligencia humana. Recordemos que Good fue un matemático británico que trabajó como criptólogo en Bletchley Park con Turing. Tras la Segunda Guerra Mundial, Good siguió tra-

bajando con Turing en el diseño de ordenadores y estadística bayesiana en la Universidad de Manchester, trasladándose luego a Estados Unidos.

Sin embargo, fue el matemático von Neumann la primera persona conocida que utilizó el concepto de «singularidad» en el contexto tecnológico. Sobre von Neumann y la computación, remitimos al excelente libro *Maniac*, de Benjamín Labatut (2023).

El concepto y el término «singularidad» fueron popularizados por Vernon Vinge (Waukesha, 1944 – La Jolla -EE.UU-, 2024), primero en 1983 en un artículo y, más tarde, en su ensayo de 1993, *The Coming Technological Singularity* (*La inminente singularidad tecnológica*) en el que escribió que señalaría el fin de la era humana, ya que la nueva superinteligencia seguiría actualizándose y avanzaría tecnológicamente a un ritmo incomprensible.



Vernon Vinge (5 may. 2006).

En palabras del propio Vinge:

> *«¿Qué es la Singularidad?*
>
> *La aceleración del progreso tecnológico ha sido el rasgo de este siglo. En este documento sostengo que estamos al borde de un cambio comparable al surgimiento de la vida humana en la Tierra. La causa de este cambio es la inminente creación por la tecnología de entidades con inteligencia superior a la humana. Existen varios medios por los que la ciencia puede lograr este avance (y esta es otra razón para tener confianza en que el acontecimiento se producirá):*
>
> *– El desarrollo de ordenadores 'despiertos' e inteligencia sobrehumana. (Hasta la fecha, la mayor controversia de la IA se refiere a si podemos crear una equivalencia humana en una máquina. Pero si la respuesta es 'sí, podemos', entonces no hay pocas dudas de que poco después se podrán construir seres.*
>
> *– Las grandes redes informáticas (y sus usuarios asociados) pueden 'despertar' como una entidad inteligente sobrehumana.*
>
> *– Las interfaces ordenador/humano pueden llegar a ser tan íntimas que los usuarios puedan considerarse razonablemente sobrehumanamente inteligentes.*
>
> *– La ciencia biológica puede encontrar formas de mejorar el intelecto humano natural.»* (Vinge, 1993)

Y concluye:

> *«De hecho, creo que la nueva era es simplemente demasiado diferente para encajar en el marco clásico del bien y el mal. Ese marco se basa en la idea de mentes aisladas e inmutables conectadas por vínculos tenues y de escaso ancho de banda. Pero el mundo posterior a la Singularidad sí encaja en la tradición más amplia de cambio y cooperación que comenzó hace mucho tiempo (quizá incluso antes del surgimiento de la vida biológica). Creo que ciertas nociones de ética se aplicarían en una era así. La investigación sobre la IA y las comunicaciones de gran ancho de banda deberían mejorar esta comprensión. Veo sólo los atisbos de esto ahora; tal vez haya reglas para distinguir el yo de los demás en función del ancho de banda de la conexión. Y aunque la mente y el yo serán mucho más lábiles que en el pasado, gran parte de lo que valoramos (conocimiento, memoria, pensamiento) no tiene por qué perderse nunca. Creo que Freeman Dyson tiene razón cuando dice que: 'Dios es aquello en lo que se convierte la mente cuando ha superado la escala de nuestra comprensión'.»* (Vinge, 1993)

Mientras no se produce la singularidad, veamos lo que hemos sido capaces los humanos de idear con nuestra imaginación. En el cine, además de muchas obras menores, las dos grandes son sin duda Terminator y Matrix.

Terminator es una serie de entregas de la película original de 1984 (véase la figura 16). La primera película fue dirigida por James Cameron, coescrita entre Cameron, Gale Anne Hurd y William Wisher Jr. y protagonizada por Arnold Schwarzenegger, Linda Hamilton y Michael Biehn. El argumento se basa en el desarrollo de un microchip capaz de dotar de inteligencia artificial a robots que luego se rebelan contra la humanidad. Se trata de una de las películas más populares sobre una hipotética guerra entre humanos y robots inteligentes capaces de crearse a sí mismos. La forma de luchar es tratar de cambiar el pasado enviando a los terminators (exterminadores) para que aniquilen a los posibles humanos que en la futura guerra liderarán el bando humano. Es muy interesan-

te ver como Arnold Schwarzenegger protagoniza a un terminator malo y a otro bueno.

Figura 16. Cartel anunciador de *The Terminator*. Cortesía de Hemdale Film Corporation©.

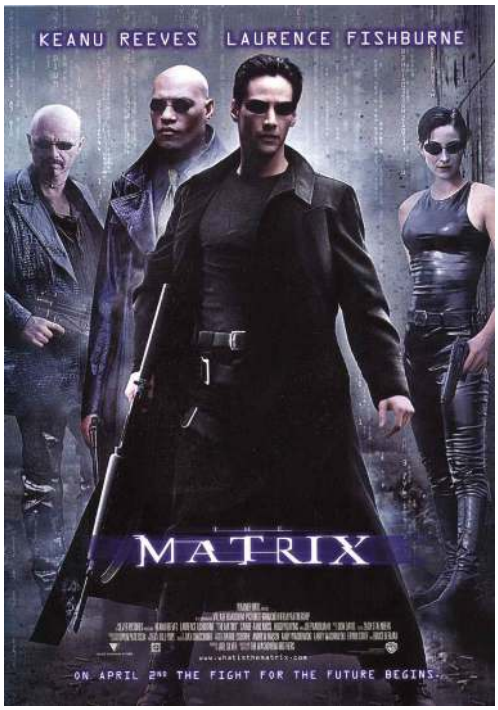


Figura 17. Cartel anunciador de *The Matrix*. Cortesía de Village Roadshow Pictures©.

Matrix es otra serie de películas iniciada en 1999, escritas y dirigidas por las hermanas Wachowski (véase la figura 17). Se compone de *The Matrix* (1999), *The Matrix Reloaded* (2003), *The Matrix Revolutions* (2003) y *The Matrix Resurrections* (2021) y están protagonizadas en sus papeles principales por Keanu Reeves, Laurence Fishburne, Carrie-Anne Moss y Hugo Weaving. En la primera película de la serie, el protagonista, Keanu Reeves interpreta a Thomas Anderson / Neo, un programador de día y hacker de noche que trata de desentrañar la verdad oculta tras una simulación conocida como «Matrix». Esta realidad simulada es producto de programas de inteligencia artificial que terminan esclavizando a la humanidad y utilizando sus cuerpos como fuente de energía.

ChatGPT y demás IAs

Llegados hasta aquí, estaría bien hacer un resumen de los conceptos que hemos ido introduciendo en las secciones previas:

La IA puede definirse como un programa capaz de sentir, razonar, actuar y adaptarse. Sin embargo, estamos todavía muy lejos de la inteligencia artificial fuerte, una máquina tan inteligente como un humano.

La inteligencia artificial se suele clasificar en dos tipos: IA general (fuerte) e IA débil (restringida).

La *IA general* es aquella con la capacidad de entender, aprender y aplicar su inteligencia a cualquier problema, de forma similar a como lo hacemos los seres humanos. Como hemos comentado, esta clase de IA no existe en la actualidad, pero es el objetivo a largo plazo y el que se quería lograr inicialmente.

La *IA débil* es aquella cuyos sistemas han sido diseñados y entrenados para realizar una tarea específica. Actualmente disponemos de esta IA, por ejemplo en asistentes virtuales como Siri o Alexa, o en diferentes aspectos de la vida cotidiana: compras en línea y otros.

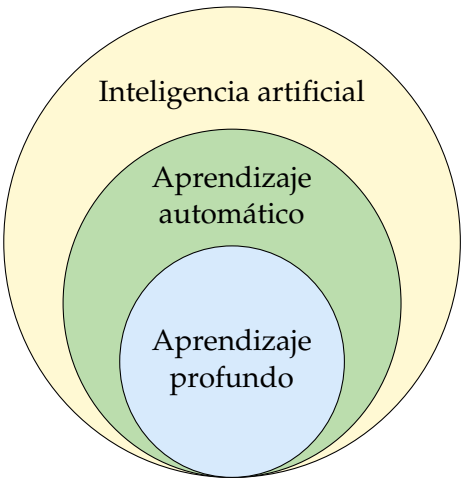


Figura 18. Jerarquías de la IA.

El *machine learning* o *aprendizaje automático* es una subcategoría de la IA débil que mejora su rendimiento a medida que procesa más datos. Un modelo de Machine Learning se entrena con datos y, cuantos más datos consume, mejores resultados puede ofrecer.

Dentro del machine learning encontramos el *deep learning* o *aprendizaje profundo*. Estos algoritmos se basan en redes neuronales que imitan el funcionamiento del cerebro humano y mejoran su rendimiento con grandes cantidades de datos.

Hay otros dos conceptos que nos conducen hacia el ChatGPT y son los siguientes.

El *Procesamiento del Lenguaje Natural* (del inglés: *Natural Language Processing*) fusiona la ciencia de la computación y el lenguaje humano. esta tecnología permite que las computadoras interpreten, manipulen y comprendan el lenguaje humano de forma efectiva.

Y los *Modelos de Lenguaje de Gran Escala* (LLM), modelos de machine learning diseñados para entender y generar texto de manera coherente y contextual. ChatGPT, en su versión más reciente (GPT-4), es un ejemplo. No es el único, hay otras empresas como Google o Facebook desarrollando sus propios LLMs. Un LLM se entrena para predecir la siguiente palabra en una frase. Esto significa que no comprende el texto como lo haría un humano, sino que genera respuestas basadas en patrones de datos a los que ha sido expuesto. Para ello, usa nociones de teoría de probabilidades.

Estos instrumentos caen dentro de lo que se ha dado en llamar la *IA generativa*, y dan respuestas que, en principio, son indistinguibles de una respuesta humana pero que contienen generalmente muchos errores, de ahí la advertencia a su uso controlado y vigilante.

La IA y la educación matemática

Esta sección fue publicada como un artículo en el Boletín de la RSME (De León *et al.*, 2024, veáse también De León y Trujillo-González, 2024); ésta es una versión ampliada y actualizada.

La Oficina C fue creada para proporcionar al Congreso de los Diputados la evidencia científica sobre temas de interés y facilitar así el diálogo entre la comunidad científica y los diputados y diputadas. Se trata pues de que los políticos tomen sus decisiones basadas en el conocimiento científico. La Oficina C está coordinada por FECYT y la Secretaría General del Congreso de los Diputados.

Desde su puesta en marcha en 2022, la Oficina ha comenzado a elaborar varios informes sobre temas de gran actualidad. Nos queremos centrar en este artículo en el reciente informe *Inteligencia artificial y educación. Retos y oportunidades en España*, publicado el pasado 29 de octubre (Calvo *et al.*, 2024). Sin duda alguna, un tema de mucho interés para la comunidad matemática española.

Una de las primeras conclusiones de este informe es la siguiente:

> *«Su uso en educación conlleva oportunidades y riesgos sobre los que todavía hay incertidumbres: aún no hay suficiente evidencia sólida e independiente sobre los efectos de introducirla en la educación y su eficacia para mejorar el aprendizaje.»*

Probablemente, no es sólo que no haya evidencia sólida, sino que más bien no se han hecho estudios sobre el tema. Lo que es preocupante, porque, según el mismo informe, sin embargo,

> *«(. . .) el 82 % de los alumnos españoles de entre 14 y 17 años, el 73 % del profesorado y el 69 % de los padres y madres encuestados dicen haber usado alguna vez alguna herramienta de IA, principalmente chatbots o asistentes virtuales. De hecho, el 40 % del alumnado afirma usar ChatGPT de manera frecuente.»*

Una vez más, la tecnología nos ha adelantado sin que estuviésemos preparados para usarla adecuadamente.

¿Y cómo se puede utilizar la IA en el ámbito educativo? El informe sugiere:

> *«La IA generativa puede apoyar al profesorado y al personal educativo en tareas como elaborar planes de sus asignaturas, generar rúbricas, adaptar textos o automatizar tareas administrativas rutinarias (planificar horarios, asignar espacios etc.). El profesorado también puede usar herramientas de IA como apoyo en las correcciones o para obtener información sobre cómo aprende el alumnado.*
>
> *En resumen, entre estas dos facetas que desde sus comienzos se atribuyen a la IA, mecanismos que sean capaces de desarrollar tareas rutinarias liberando de esos quehaceres a los humanos (convivimos con esta IA desde hace décadas) o mecanismos que son capaces de pensar por sí mismos y proponer nuevas ideas, nos quedamos con las primeras, las rutinarias. Pero el*

profesorado puede ir un poco más allá.»

Sin embargo, la IA puede ser utilizada en el aula de matemáticas de una manera mucho más interesante y estimulante, y no nos referimos simplemente al desarrollo de las áreas de pensamiento computacional, el lenguaje de programación y la robótica educativa, ya implementadas hace años desde infantil, primaria y secundaria.

En primer lugar, ChatGPT nos responde porque es capaz de asociar una palabra detrás de otra a una velocidad tan grande que sus respuestas son automáticas, y nos da la sensación de estar hablando realmente con alguien inteligente. Pero esta asociación casi markoviana nos da pie para enseñar a nuestros alumnos muchos conceptos de la estadística (procesos bayesianos) y la teoría de probabilidades. Y no sólo de la estadística, sino de matrices, espacios vectoriales o teoría de grafos (redes neuronales). Todo ello está dentro de ChatGPT. Así como usar sus múltiples errores de salida para potenciar el rigor, la revisión crítica de cualquier supuesto (venga de donde venga) y, a su vez, concienciar a los estudiantes de matemáticas de que de los errores en el desarrollo de las matemáticas también se puede aprender.

Pero estas enseñanzas trascienden a las matemáticas. Procesar estas preguntas con la información disponible en las bases de datos y con semejante rapidez requiere grandes infraestructuras de cómputo. Cuando hablamos de la nube, olvidamos que ésta no está en el cielo, sino que se materializa en gigantescas granjas de ordenadores. Los grandes modelos lingüísticos (LLM) como ChatGPT son algunas de las tecnologías que más energía consumen (de hecho, se espera que el 30 % de la energía consumida por la humanidad en 2050 sea debido a la computación). Y como cada pregunta se procesa a través de un servidor que ejecuta miles de cálculos en los centros de datos, éstos generan calor, y se precisa de sistemas de agua para refrigerar el equipo y mantenerlo en funcionamiento. Por lo tanto, tenemos un consumo energético enorme y un consumo de agua insostenible. Sólo para refrigerar las máquinas que entrenaron a ChatGPT-3 se usaron unos 700 000 litros de agua. Para generar un correo electrónico de 100 palabras, ChatGPT-4 emplea un poco más de una botella de agua. Estas ideas se pueden transmitir también a nuestros alumnos para que sean conscientes del problema de sostenibilidad que afrontamos en los próximos años (Green AI), que ya tenemos por el uso de los móviles o la basura tecnológica que está generando esta explosión de la IA.

Y también se puede utilizar para evidenciar y combatir sesgos, ya que la IA reproduce los propios de los humanos, como se ha visto en los textos e imágenes producidos por estos modelos que llevan a generar contenido sexista, racista o a perpetuar y replicar determinados estereotipos. De nuevo, apuntamos nuestra mirada al fomento del espíritu crítico, la preocupación por los detalles y el análisis sin prejuicios, elementos transversales para el desarrollo profesional de cualquier ciudadano.

Mientras la legislación española y las transposiciones de las regulaciones europeas no se hayan introducido en nuestro sistema educativo, nos conviene ir ya caminando en la buena dirección, seguramente de la mano de la IA. Hasta ahora, podemos ya consultar algunas fuentes relevantes sobre estos temas: (Miao *et al.*, 2021), un informe global de la UNESCO y (Tuomi, 2018), un informe para el ámbito europeo.

Hay dos comunidades autónomas, Cataluña y Canarias, que ya han elaborado una guía al respecto siguiendo estas directrices que vienen de Europa y la UNESCO. Por su interés, las citamos a continuación: *La inteligencia artificial en el ámbito educativo* (Área de Tecnología Educativa de la Dirección General de Ordenación de las Enseñanzas, Inclusión e Innovación, 2024) y *La inteligència artificial en l'educació. Orientacions i recomanacions per al seu ús als centres* (Direcció General d'Innovació, Digitalització i Currículum, 2024). Esperamos que cunda el ejemplo y vayan apareciendo nuevos informes y que haya una buena coordinación en todas las políticas educativas relativas a la IA.

La IA en la investigación matemática

Si nos referíamos antes a la necesidad de reflexión en cómo la IA debe ser considerada en los procesos educativos y en cómo puede ayudar en los mismos, no es menos relevante el papel que la IA ha ido desempeñando en la propia investigación matemática. Un reciente artículo de Terence Tao (2025) incide sobre estos aspectos.

Usar instrumentos para facilitar sus cálculos ha sido siempre una ocupación de los matemáticos, culminada con la invención de los ordenadores y programas que son capaces no sólo de dar aproximaciones numéricas (MatLab©), sino también de realizar cálculos simbólicos (Mathematica©). Es más, con los ordenadores se han podido desarrollar las llamadas demostraciones asistidas por ordenador. Se usa algún programa para comprobar todos los casos posibles, lo

que a mano no seríamos capaces de decidir en años. Dos ejemplos paradigmáticos han sido la prueba en 1976 del teorema de los cuatro colores («*dado cualquier mapa geográfico con regiones continuas, éste puede ser coloreado con cuatro colores diferentes o menos, de forma que no queden regiones adyacentes con el mismo color*») y la conjetura de Kepler («*si apilamos esferas iguales, la densidad máxima se alcanza con una apilamiento piramidal de caras centradas*»). El primer problema fue resuelto por Kenneth Appel (Nueva York, 1932 – Dover -EE.UU-, 2013) y Wolfgang Haken (Berlín, 1928 – Champaign -EE.UU-, 2022), mientras que el segundo se debe a Thomas C. Hales (San Antonio, 1958), en 1998. En este último caso, la comunidad matemática expresó dudas sobre la validez de un teorema demostrado usando un programa de ordenador. Como señala Bonet (2024):

> «*Las principales objeciones a este tipo de pruebas son la posibilidad de errores en la programación y, sobre todo, la gran dificultad de entender y verificar estas demostraciones obtenidas por el ordenador.*»

En su artículo, Tao sugiere una serie de aplicaciones que ayudarían en la investigación matemática y que reproducimos textualmente:

> «*– Los algoritmos de aprendizaje automático pueden utilizarse para descubrir nuevas relaciones matemáticas o generar ejemplos potenciales o contraejemplos de problemas matemáticos.*
>
> *– Los asistentes de pruebas formales pueden utilizarse para verificar pruebas (así como la salida de grandes modelos de lenguaje), permitir colaboraciones matemáticas a gran escala y ayudar a construir conjuntos de datos para entrenar a los algoritmos de aprendizaje automático antes mencionados.*
>
> *– Los modelos de lenguaje de gran escala, como ChatGPT, pueden utilizarse (potencialmente) para facilitar y agilizar el uso de otras herramientas; también pueden sugerir estrategias de demostración o trabajos relacionados, e incluso generar pruebas (sencillas) de forma directa.*» (Tao, 2025)

Se avecina un futuro con tiempos interesantes, pero no está de más recordar aquí esta reflexión de Godfrey H. Hardy (Cranleigh, 1877 – Cambridge, 1947). Cuando en 1928 David Hilbert planteó el siguiente problema: «*encontrar un procedimiento general que permita probar o refutar cualquier proposición matemática, es decir, mecanizar todas las matemáticas*», Hardy escribió:

> «*Supongamos, por ejemplo, que pudiéramos encontrar un sistema finito de reglas que nos permitiera decir si una fórmula dada es demostrable o no. Este sistema constituiría un teorema metamatemático. Por supuesto, no existe tal teorema, lo cual es muy afortunado, ya que si existiera tendríamos un conjunto mecánico de reglas para la solución de todos los problemas matemáticos, y nuestras actividades como matemáticos llegarían a su fin.*» (Hardy, 1929)

A modo de conclusiones

A modo de conclusión, nos enfrentamos a un nuevo paradigma, que por una parte ofrece grandes oportunidades, pero por otra despierta nuestros miedos. Pero como hemos visto, la IA, de una manera más o menos desarrollada, ha convivido con la humanidad y hemos podido mejorar nuestras vidas con ella. Y éste es el mensaje con el que debemos enfrentar las próximas décadas.

Con cierta precaución, ya que son las grandes compañías las que tienen la capacidad de desarrollar potentes IAs así como algunos gobiernos que todavía no ofrecen las garantías democráticas de los países de Occidente, las cuestiones éticas van a ser cada vez más relevantes y la ciudadanía debe estar alerta mano a mano con los científicos.

Como dice Gabriela Ramos, Subdirectora General de Ciencias Sociales y Humanas de la UNESCO, en las conclusiones del 2º Foro Mundial sobre la Ética de la Inteligencia Artificial: Cambiando el panorama de la gobernanza de la IA, que tuvo lugar en el Centro de Congresos Brdo de Kranj (Slovenia) los días 5 y 6 de febrero de 2024:

> «*El rápido auge de la inteligencia artificial (IA) ha generado nuevas oportunidades a nivel global: desde facilitar los diagnósticos de salud hasta posibilitar las conexiones humanas a través de las redes sociales, así como aumentar la eficiencia laboral mediante la automatización de tareas.*
>
> *Sin embargo, estos rápidos cambios también plantean profundos dilemas éticos, que surgen del potencial que tienen los sistemas basados en IA para reproducir prejuicios, contribuir a la degradación del clima y amenazar los derechos humanos, entre otros. Estos riesgos asociados a la IA se suman a las desigualdades ya existentes, perjudicando aún más a grupos históricamente marginados.*
>
> *En ninguna otra especialidad necesitamos más una 'brújula ética' que en la inteligencia artificial. Estas tecnologías de utilidad general están remodelando nuestra*

forma de trabajar, interactuar y vivir. El mundo está a punto de cambiar a un ritmo que no se veía desde el despliegue de la imprenta hace más de seis siglos. La tecnología de inteligencia artificial aporta grandes beneficios en muchos ámbitos, pero sin unas barreras éticas corre el riesgo de reproducir los prejuicios y la discriminación del mundo real, alimentar las divisiones y amenazar los derechos humanos y las libertades fundamentales.»

Referencias bibliográficas

Área de Tecnología Educativa de la Dirección General de Ordenación de las Enseñanzas, Inclusión e Innovación. (2024). *La inteligencia artificial en el ámbito educativo* (inf. téc.). Consejería de Educación, Formación Profesional, Actividad Física y Deportes de Canarias. https://n9.cl/mquv1u

Asimov, I. (2022). *Yo, robot*. Edhasa.

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorrell, T., Wallis, M., Whitby, B. & Winfield, A. (2011). *Principles of robotics. Regulating robots in the real world* (inf. téc.). The Engineering y Physical Sciences Research Council (EPSRC). https://doi.org/10.1080/09540091.2016.1271400

Bonet, J. (2024). *Demostraciones matemáticas y el ordenador*. Real Academia de Ciencias de España.

Calvo, J., Cobo, C., de la Higera, C., de Salvador Carrasco, L., Díaz Rodríguez, I., Fernández, R., González González, C., Lobo Martínez, J., López Cobo, M., Mateos Caballero, A., Mesa Sánchez, C., Ortíz de Zárate, L., Alcarazo, Poyatos Dorado, C., Rodríguez García, J., Sánchez Vera, M., Sierra, C., Tarabini-Castellani Clemente, A., Vallés Peris, N. & Wasson, B. (2024). *Inteligencia artificial y educación* (inf. téc.). Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C). https://doi.org/10.57952/hqct-6d69

Copeland, B. (2013). *Alan Turing. El pionero de la era de la informática*. Turner.

De León, M. & Timón, Á. (2014). *Rompiendo códigos. Vida y legado de Turing*. Catarata.

De León, M. & Trujillo-González, R. (2024). Por qué debemos enseñar Machine Learning en todos los títulos universitarios. *Boletín de la RSME*, (864).

De León, M., Vélez-Melón, M. & Trujillo-González, R. (2024). A vueltas con la Educación y la IA. *Boletín de la RSME*, (872).

Degli-Esposti, S. (2023). *La ética de la inteligencia artificial*. Catarata.

Direcció General d'Innovació, Digitalització i Currículum. (2024). *La intel·ligència artificial en l'educació. Orientacions i recomanacions per al seu ús als centres* (inf. téc.). Departament d'Educació. Generalitat de Catalunya. https://n9.cl/egunz

Hardy, G. (1929). Mathematical proof. *Mind, New Series*, *38*(149), 1-25. https://is.gd/SSOyqs

Jevons, W. (1870). On the mechanical performance of logical inference. *Philosophical Transactions of the Royal Society*, *160*, 497-518. https://is.gd/ykdhTF

Labatut, B. (2023). *Maniac*. Anagrama.

McCarthy, J., Minsky, M., Rochester, N. & Shannon, C. (1955). *A proposal for the Dartmouth summer research project on artificial inteligence* (inf. téc.). https://n9.cl/kqiym

Miao, F., Holmes, W., Ronghuai, H. & Hui, Z. (2021). *Inteligencia artificial y educación: guía para las personas a cargo de formular políticas* (inf. téc.). UNESCO. https://is.gd/em8VCT

Mitchell, M. (2024). *Inteligencia Artificial. Guía para seres pensantes*. Capitán Swing.

Ríos-Insua, D. & Gómez-Ullate, D. (2019). *Big data. Conceptos, tecnologías y aplicaciones*. Catarata.

Ruiz-Sarrias, O. (2024). *Las matemáticas de la IA. Introducción al deep learning*. Catarata.

Tao, T. (2025). Machine-assisted proof. *Notices of the AMS*, *72*(1), 6-13. https://is.gd/7hCj24

Tuomi, I. (2018). *The impact of artificial intelligence on learning, teaching, and education* (Y. Punie, R. Vuorikari & M. Cabrera, Eds.; inf. téc.). European Commission: Joint Research Centre. https://doi.org/10.2760/12297

Turing, A. M. (2012). *¿Puede pensar una máquina?* KRK ediciones.

Vinge, V. (1993). *The coming technological singularity: How to survive in the post-human era* (inf. téc.). VISION-21 Symposium. https://n9.cl/gsp7v

Fecha de recepción: *2 de diciembre, 2024.*

Fecha de aceptación: *12 de enero, 2025.*

Manuel de León (Requejo –Zamora–, dic. 1953)

Profesor de Investigación del CSIC, se ha dedicado principalmente al estudio de la Geometría Diferencial y sus aplicaciones a la Mecánica y a la Física Matemática, campos en los que ha publicado más de 350 artículos y 5 monografías. De acuerdo con el *ranking* de Stanford de 2024, aparece en el 2 % de los científicos más destacados en el mundo.

Entre sus muchos logros y distinciones destaca el haber sido fundador y Director del Instituto de Ciencias Matemáticas (ICMAT), centro mixto de investigación entre el CSIC y las universidades UAM, UCM, y UC3M (es uno de los Centros de Excelencia en España del Programa Severo Ochoa, un galardón que se ha repetido ya en cuatro convocatorias consecutivas). Fue Vicepresidente de la Comisión Gestora que refundó la Real Sociedad Matemática Española en 1996 y miembro de su ejecutiva durante 10 años. Ha sido el promotor de la creación del Comité Español de Matemáticas (CEMat) y su primer Presidente (2004-2007). También fue Presidente del Comité Ejecutivo del ICM2006-Madrid. Ha sido miembro del Comité Ejecutivo de la Unión Matemática Internacional (IMU) durante dos mandatos (2007-2014), miembro del Comité Ejecutivo del Consejo Internacional de la Ciencia, ISC (2014 – 2018) y actualmente es miembro del Management Group de ISC-Europa. Es académico de número de la Real Academia de Ciencias Exactas, Físicas y Naturales de España de la que es su Tesorero, Director de ESTALMAT y Presidente de su Comité de Relaciones Internacionales. Es además académico correspondiente de la Real Academia Canaria de Ciencias y de la Real Academia Galega de Ciencias. Ha sido muy activo en la gestión de la ciencia como coordinador en la Agencia Nacional de Evaluación y Prospectiva ANEP (2003 – 2011), de la Comisión de área de Ciencias y Tecnologías Físicas del CSIC (2000 – 2008) y del Comité PESC de la European Science Foundation, ESF (2006 – 2011). Es autor de numerosos trabajos de divulgación de la Ciencia, de las Matemáticas en particular: Libros, contribuciones en el blog Matemáticas y sus fronteras, muchos artículos en *El País*, *The Conversation*, *El Mundo*, entre otros, y ha impartido más de un centenar de conferencias de divulgación en centros educativos, museos y centros culturales.

✉ mdeleon@icmat.es

ICMAT (CSIC) y Real Academia de Ciencias