

De la neurona de Cajal a la superinteligencia artificial: paralelismo cerebral-artificial

From Cajal's neuron doctrine to artificial superintelligence: the brain-AI parallel

Rubén F. Hernán-Gómez

Servicio de Neurología, Hospital Universitario Puerta de Hierro de Majadahonda, Madrid, España

Resumen

La historia de la inteligencia artificial (IA) se remonta a los albores de la ciencia, permaneciendo en silencio durante numerosos siglos para resurgir al inicio de la Edad Contemporánea y, finalmente, convertirse en el principal foco de atención de nuestro tiempo. Esta revisión narrativa es mucho más que un viaje por las vidas de aquellas personas que han contribuido al desarrollo de la IA moderna. Mediante un análisis en profundidad de la mutua influencia entre la neurología y las ciencias de la computación, se pone de manifiesto cómo las redes neuronales artificiales no son más que el intento del homo sapiens de recrear el mayor sistema inteligente conocido hasta el momento: el cerebro humano. Beneficios trascendentales, desafíos crecientes y riesgos inquietantes convergen en torno a la indómita estela de la Cuarta Revolución Industrial.

Palabras clave: Neurología. Redes neuronales. Inteligencia artificial. Aprendizaje automático. Aprendizaje profundo. Singularidad.

Abstract

The history of artificial intelligence (AI) dates to the dawn of science, remaining silent for centuries, only to reappear at the beginning of the Contemporary Age as a flagship of neurotechnology. This narrative review goes further than just covering a journey through the lives of those who have contributed to the development of modern AI. Through an in-depth analysis of the mutual influence that Neurology and Computer Science are subject to, it becomes clear that artificial neural networks are merely our species' attempt to recreate the greatest intelligent system known to date: the human brain. Unprecedented benefits, growing challenges and disquieting risks intertwine at the core of the Fourth Industrial Revolution.

Keywords: Neurology. Neural networks. Artificial intelligence. Machine learning. Deep learning. Singularity.

Correspondencia:

Rubén F. Hernán-Gómez

E-mail: rubenhernan.neuro@gmail.com

1577-8843 / © 2025. Kranion. Publicado por Permanyer. Este es un artículo open access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Fecha de recepción: 12-03-2025

Fecha de aceptación: 19-03-2025

DOI: 10.24875/KRANION.M25000096

Disponible en internet: 14-07-2025

Kranion. 2025;20(1):37-47

www.kranion.es

Introducción

Hubo una época en la que toda computadora era humana y ninguna neurona podía ser calificada como artificial¹. Un breve periodo marcado por la doctrina del reticularismo y el rechazo a la teoría celular². Décadas prolíficas en las que aquellos que sentaban las bases de la computación ejercían también la neurociencia. Un tiempo en que la amenaza de una inteligencia superior incorpórea no ocupaba portadas de diarios. Un último instante de ingenuidad en que creímos que ningún ordenador podría hacernos frente en el milenarismo arte de la estrategia de tablero³. El siglo XXI es una vorágine de cambio constante vertebrada por un mundo físico cada vez más poblado por máquinas y un mundo digital cada vez más absorbente para los humanos⁴. En medio de la inversión del espacio de existencia, surgen inevitables interrogantes: ¿qué es la inteligencia? ¿puede lo artificial ser inteligente? ¿somos acaso tan diferentes de nuestra propia creación?

Carbono y silicio

El cerebro es el sistema inteligente no artificial más complejo y eficiente conocido hasta la fecha, con una estructura basada en el carbono⁵. «Inteligencia» es un concepto abstracto, pero podría definirse como un elemento que surge de la coordinación de diversas funciones cerebrales y que se expresa por medio de sofisticados comportamientos con la finalidad de superar las dificultades que plantea un entorno externo cambiante. Extender este término más allá de los organismos biológicos requiere, al menos, del cumplimiento del siguiente criterio: poseer capacidad para aprender nuevas nociones que permitan adaptarse a un ambiente inestable, asumiendo que sus recursos son limitados⁶. «Artificial» es un adjetivo habitualmente peyorativo: opuesto a lo natural, lo biológico o lo veraz⁶. Resulta especialmente paradójico que su primera acepción en el Diccionario de la Lengua Española sea la más olvidada: aquello que ha sido producido por el ser humano.

La inteligencia artificial (IA) es la rama de las ciencias de la computación que pretende emular la cognición humana, así como el conjunto de modelos computacionales que exhiben dicho intelecto. No se trata de una única entidad, sino de un amplio espectro de herramientas digitales contenidas en chips de silicio⁷.

Histor[IA] de dos neuronas

Los principios matemáticos y lógicos de las grandes civilizaciones antiguas (sumeria, egipcia, griega, árabe...)

son la chispa que encendería, hace miles de años, la mecha del irregular pero creciente proceso de desarrollo de la IA. Ha transcurrido mucho tiempo desde los primeros vestigios de computación analógica (destacando por su complejidad el mecanismo de Anticitera), o desde la primera concepción de máquinas pensantes hasta el desarrollo acelerado de la IA contemporánea con la llegada de la Cuarta Revolución Industrial (a partir de 2010)^{8,9}; a continuación se plasma un recorrido por los eventos más destacables de la historia de tan vanguardista tecnología por medio de aquellas personas que impulsaron su evolución. No se debe perder de vista que el progreso científico acontece en paralelo en todo el planeta, por lo que ciertas atribuciones podrían estar sujetas a discusión.

Ramón Llull (1232-1316)

Nacido en Palma de Mallorca, fue un filósofo, teólogo, misionero y escritor de numerosas obras de corte espiritual. En su tratado *Ars Magna* describió el arte combinatorio, una técnica semejable a un modelo generativo primitivo, con la finalidad de dar sentido matemático al razonamiento humano. Construyó, en términos relativos, una máquina lógica: un sistema de discos rotatorios con símbolos grabados, capaz de producir diferentes enunciados que verificar o refutar utilizando el propio dispositivo. Resulta paradójico pensar que, siendo su objetivo demostrar mediante las combinaciones que el cristianismo era la «religión verdadera», su invento haría de él uno de los precursores de las ciencias de la computación¹⁰.

Charles Babbage (1791-1871)

Matemático británico considerado como creador del primer diseño (en papel) de una computadora moderna, en 1822, con la intención de que resolviese complejos cálculos matemáticos. Para ello se basó en el telar mecánico de Jacquard, que utilizaba tarjetas perforadas como instrucciones para tejer. Si bien no llegó a ver su ordenador construido en vida, tras su muerte se demostró que era realmente funcional. Trabajó estrechamente con Ada Lovelace en este proyecto¹¹.

Ada Lovelace (1815-1852)

Augusta Ada King, condesa de Lovelace (hija de Lord Byron), fue una brillante matemática y escritora británica, que supo ver más allá que el propio Charles Babbage en lo referente a computadoras. Su fama procede principalmente de la traducción y ampliación del trabajo que él inició, siendo la responsable de redactar

el primer programa informático diseñado para ejecutarse en un ordenador. Predijo limitaciones de la IA a largo plazo, aún vigentes a día de hoy¹¹.

Santiago Ramón y Cajal (1852-1934)

Padre de la neurociencia moderna y polímata en el sentido más amplio de la palabra (médico, histólogo, científico, inventor, dibujante, escritor, profesor...), galardonado con el premio Nobel de Medicina en 1906¹². Demostró la individualidad de la neurona como célula e hizo de ella la unidad anatómica y funcional principal del sistema nervioso (doctrina neuronal, en oposición a la doctrina reticular de von Gerlach y Golgi), describió la direccionalidad que siguen los impulsos nerviosos, introdujo la idea de plasticidad neuronal y realizó excepcionales descripciones gráficas de neuroanatomía (cartografía del sistema nervioso tanto a nivel central como periférico). Cabe destacar que la palabra «neurona» no fue acuñada por el propio Cajal (se atribuye a Wilhem Waldeyer en 1881), que se refería a ellas con varios términos sinónimos¹³.

Warren McCulloch (1898-1969)

Neurocientífico y psiquiatra estadounidense coautor del modelo neuronal de McCulloch-Pitts (inspirado en los descubrimientos de Cajal y la circuitería reverberante de Lorente de Nó)², que pretende simular de forma digital una neurona biológica mediante algoritmos que funcionan emulando la «ley del todo o nada» propia del potencial de acción. Sentó las bases de las futuras redes neuronales artificiales (ANN, del inglés *artificial neural networks*). Colaboró con figuras como von Neumann en el desarrollo de teorías sobre control de sistemas biológicos, electrónicos y *feedback*¹⁴.

Rafael Lorente de Nó (1902-1990)

Neurólogo y neurofisiólogo español, discípulo directo de Ramón y Cajal, conocido por proponer el modelo no lineal del córtex cerebral, basado en complejos circuitos neuronales con conexiones recurrentes. Demostró la existencia de bucles de retroalimentación en el tronco del encéfalo (aunque el estímulo nervioso siempre viaje en el mismo sentido, puede transmitirse hasta regiones proximales a su localización actual mediante estos circuitos). Concebía la neurona como un dispositivo de suma- ción spatiotemporal que producía respuestas en forma de potenciales de acción. Sus contribuciones serían decisivas para el florecimiento de la computación moderna².

John von Neumann (1903-1957)

Matemático, informático y físico de origen húngaro. Famoso por sus aportaciones en materia de armas nucleares (bomba atómica y bomba de hidrógeno) derivadas de la creación de su «arquitectura de von Neumann», modelo en que se basan sus computadoras MANIAC y la mayoría de los ordenadores actuales (caracterizado por una unidad de procesamiento o CPU, una unidad de memoria y señales de entrada/salida). Su interés en formas biológicas de procesamiento de información provino del modelo de McCulloch-Pitts, llegando a concebir el cerebro con una gran computadora digital (que inspira directamente la arquitectura mencionada previamente). Formaba parte de un grupo académico que investigaba cómo replicar comportamiento humano mediante procedimientos electrónicos. En lo respectivo a máquinas inteligentes, pretendía construir un autó- mata artificial con capacidad de replicarse^{2,15}.

Donald Hebb (1904-1985)

Psicólogo y neurocientífico canadiense que introdujo la teoría del aprendizaje hebbiano: *neurons that fire together wire together* (las neuronas que se activan de forma conjunta se conectan). Esta concepción se encuentra en la base de la plasticidad cerebral y es clave en la formación de la memoria^{2,3}.

Alan Turing (1912-1954)

Matemático británico considerado fundador de las ciencias de la computación y de la rama de la IA. En 1936 presentó su modelo teórico de computadora programable capaz de ejecutar todo tipo de algoritmos, si bien era consciente de que las limitaciones técnicas de la época hacían imposible que se materializase a corto plazo (años tras su fallecimiento se demostró que su máquina funcionaba como él había previsto). Su intención final era construir ordenadores capaces de imitar al cerebro humano, fiel creyente en que podrían llegar a pensar y razonar (para ello se basó en la estimación de capacidad de memoria cerebral)¹¹. Creador del test que lleva su nombre como medida de la inteligencia de una computadora y colaborador durante un breve periodo de tiempo con von Neumann, aunque posteriormente trabajarían de forma completamente independiente y bajo ideales disímiles. Con gran acierto describió como prometedores campos de aplicación de IA los siguientes: ajedrez y juegos de cartas, aprendizaje de idiomas, traducción, criptografía y matemáticas¹⁵.

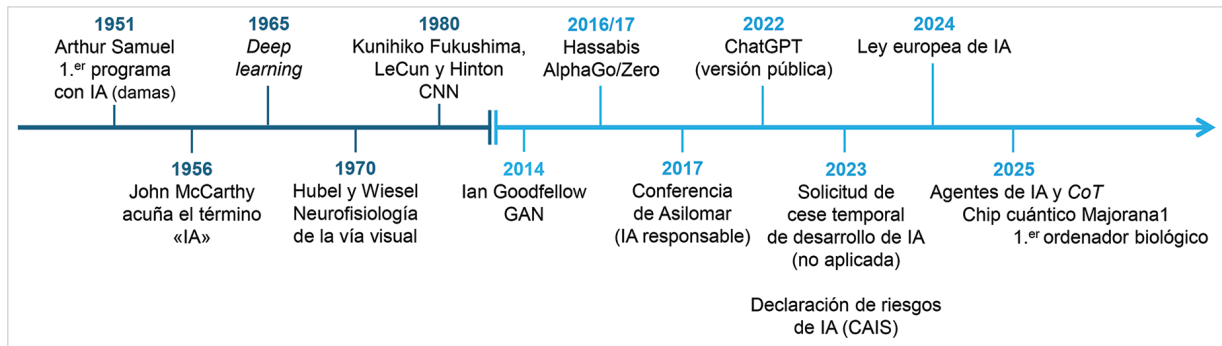


Figura 1. Historia de la inteligencia artificial desde la segunda mitad del siglo XX. CAIS: *Center for AI Safety*; CNN: red neuronal convolucional; CoT: razonamiento secuencial tipo *Chain-of-Thought*; GAN: red generativa antagonista.

Walter Pitts (1923-1969)

Matemático y lógico autodidacta de nacionalidad estadounidense. Su colaboración con McCulloch desembocó en el primer modelo conceptual de una red neuronal artificial. No se hacía explícito el funcionamiento interno de cada neurona, pero sí el esquema de conexiones que permitiría que funcionasen como operadores lógicos^{2,11,14}.

Frank Rosenblatt (1928-1971)

Psicólogo e informático teórico estadounidense que diseñó una de las primeras ANN, denominada «perceptrón», capaz de reconocer relaciones en la información de entrada y aprender de sus errores. Aunque hipotetizó los perceptrones multicapa (que derivarían en el origen del *deep learning* [DL]), su perceptrón representa realmente una única neurona digital. Las duras críticas recibidas tras presentar su trabajo contribuyeron al «invierno de la IA de 1970», una época de desaceleración del desarrollo de esta tecnología¹⁴.

Geoffrey Hinton (1947-)

Científico computacional y psicólogo cognitivo británico. Junto con John Hopfield recibió el premio Nobel de Física en 2024 por sus aportaciones a la tecnología de *machine learning* (ML). Se le ha atribuido en varias ocasiones el título de creador del algoritmo de retropropagación (que se encuentra en la base del aprendizaje de la IA moderna), cuestión que él mismo ha desmentido. Las investigaciones de ambos científicos permitieron optimizar las ANN en áreas como la capacidad de memoria, el análisis masivo de datos y la habilidad para descubrir patrones en estos de forma autónoma^{16,17}.

Yann Lecun (1960-)

Ingeniero de *software* e informático teórico especializado en aprendizaje profundo, de origen francés. Sus aportaciones han permitido potenciar las capacidades de las redes neuronales convolucionales, especialmente en el campo de la identificación de caracteres escritos a mano, reconocimiento de objetos cotidianos y de caras humanas. Ha trabajado en prevención de desinformación en redes sociales mediante detección automatizada por IA¹⁸.

Demis Hassabis (1976-)

Ingeniero de IA británico, informático teórico y científico de datos. Creador de la compañía DeepMind (actualmente propiedad de Google) y diseñador de AlphaZero, un programa capaz de aprender de forma autónoma a jugar al ajedrez, al *go* y al *shogi* a un nivel insuperable por el ser humano u otros programas. Galardonado con el premio Nobel de Química en 2024 por el desarrollo de AlphaFold2, un modelo de IA capaz de predecir con exactitud la forma tridimensional de una proteína con base en la secuencia de aminoácidos que la componen, siendo numerosas sus potenciales aplicaciones en investigación científica¹⁶.

Hacia la Cuarta Revolución Industrial

Desde la última mitad del siglo XX la historia de la IA se torna más enrevesada y difusa, con un número muy elevado de contribuyentes a su progreso. La **figura 1** complementa los apartados previos con una relación de acontecimientos relevantes hasta llegar a la actualidad^{4,11}.

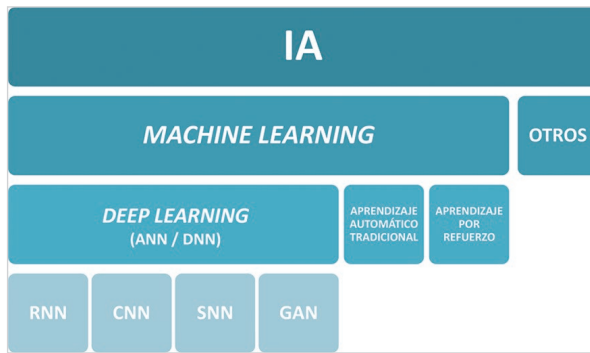


Figura 2. Jerarquía simplificada de la inteligencia artificial contemporánea.

¿Puede una máquina pensar?

Con esta pregunta comienza el famoso ensayo de Alan Turing, *Computing machinery and intelligence*¹⁵. Como se ha mencionado previamente, la neurociencia y la IA han compartido históricamente un fuerte vínculo bidireccional¹⁹. La arquitectura cerebral es el gran plano en el que se inspiran los modelos de IA contemporánea, pero tratar de imitar el funcionamiento de mil billones de sinapsis es uno de los mayores retos a los que nos hayamos enfrentado como especie, más si cabe cuando nuestra comprensión sobre el funcionamiento del cerebro humano es limitada²⁰. Se denomina aprendizaje automático o ML a la disciplina de IA que hace posible que un sistema informático aprenda a partir de los datos que procesa, sin estar dicho aprendizaje explícitamente programado. El aprendizaje profundo o DL es el subcampo de ML que utiliza ANN para el reconocimiento de representaciones complejas en los datos^{21,22}. La **figura 2** ofrece una visión holística de la IA actual.

Neurofisiología artificial

Profundicemos en el símil anatómico y funcional existente entre lo biológico y lo artificial. Para facilitar la utilización de un lenguaje de investigación común, aquellas siglas directamente relacionadas con ingeniería informática se transcriben según su voz anglosajona.

Neurona, perceptrón, sinapsis

De igual forma que la neurona biológica es la unidad fundamental del sistema nervioso, la neurona artificial es el componente clave en las arquitecturas de aprendizaje profundo. El perceptrón de Rosenblatt fue uno de los primeros ejemplos de neurona artificial¹⁴; la

figura 3 realiza una comparación ilustrativa con su contrapartida orgánica.

Redes neuronales

Las neuronas humanas se agrupan en complejas redes biológicas (BNN, del inglés *biological neural networks*) que se extienden hacia todos los confines del sistema nervioso. A su vez, las neuronas digitales forman parte de ANN, denominadas profundas (DNN, del inglés *deep neural networks*) si contienen un número elevado de capas ocultas²³. Los perceptrones multicapa son la versión más simple de ANN²². Gracias a las ANN no es necesario programar explícitamente instrucciones para que un ordenador las ejecute, sino que se entrena al propio programa para que encuentre las conexiones deseadas entre sus neuronas, siendo capaz de resolver autónomamente el problema planteado³. Descifrar qué sucede exactamente en el interior de una DNN reviste gran complejidad, pues sus parámetros se autorregulan constantemente²³.

Atención

En el ser humano, la atención depende de la interrelación de circuitos neuronales subespecializados a nivel del sistema nervioso central (incluyendo diversas regiones del córtex cerebral, ganglios basales, diencefalo y troncoencéfalo), con cierta participación del sistema nervioso autónomo²⁴. Es imprescindible para concentrarse en determinados estímulos e ignorar otros⁷. Los modelos artificiales intentan aproximarse a este concepto fragmentando la información de entrada y analizando qué valor aporta cada fracción al contexto global. Esto es posible mediante la arquitectura de transformadores (*transformers*), que asigna distinta relevancia a los elementos que componen el *input* según las relaciones de dependencia que establecen entre sí^{20,25}. Los transformadores optimizan la capacidad de procesamiento de texto o imagen de las ANN, emulando mecanismos biológicos de atención visual, espacial o interna, a expensas de un coste elevado^{7,19}.

Memoria

Es la capacidad para evocar información adquirida previamente, esencial para el aprendizaje y la supervivencia, pues media muchas otras habilidades cognitivas. Depende de redes neuronales de distribución difusa, con especial implicación de las distintas áreas corticales sensorio-perceptivas y del sistema límbico^{19,20}.

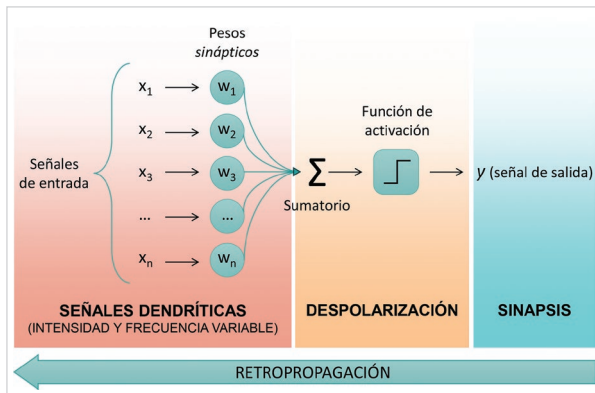


Figura 3. Perceptrón y analogía con la neurona biológica. Este tipo de neurona artificial actúa ponderando las múltiples entradas que recibe, calculando su suma y aplicando una función matemática de cuyo resultado depende que se emita una señal hacia la siguiente neurona o permanezca inactiva. Puede «aprender» mediante técnicas como la retropropagación.

Las redes neuronales recurrentes (RNN, del inglés *recurrent neural networks*) utilizan conexiones repetitivas para recuperar cierta parte de los datos procesados con el paso del tiempo. Son ideales para tareas de traducción, al lograr mantener y reintegrar información desde la memoria recuperable²⁰. La memoria humana puede tener carácter inmediato, a corto, medio o largo plazo; en las ANN sucede algo parecido desde el surgimiento de los LSTM (*long short-term memory models*), que pueden mantener inalterados ciertos datos durante un breve espacio de tiempo (memoria de trabajo) o tras miles de ciclos de entrenamiento (largo plazo). Del mismo modo que la consolidación de la memoria requiere de la repetición del contenido recientemente adquirido a nivel del hipocampo y el neocórtex, algunas herramientas de IA cuentan con una función de reexposición (*replay*) a eventos virtuales que les permiten alcanzar un mayor grado de desempeño en una tarea¹⁹. Los recursos de una RNN, como los del encéfalo, son limitados, por lo que ambos están sujetos al olvido de contenido si se sobrepasa el punto de inflexión⁶.

Lenguaje

Tradicionalmente considerada como una actividad cognitiva específicamente humana, es la gran ventaja evolutiva de nuestra especie. El procesamiento del lenguaje depende de intrincadas conexiones cerebrales, más allá del modelo clásico de Broca y Wernicke,

y del apoyo en otras funciones superiores⁷. Los grandes modelos de lenguaje (LLM, del inglés *large language models*) son sistemas de IA capaces de procesar y generar lenguaje natural, ocasionalmente alojados en el seno de arquitecturas multimodales más grandes (que permiten *input* visual o de audio)²⁶. Buscan la lógica profunda que subyace a la lengua, pero prescinden absolutamente de la semántica. Han sido expuestos a tal cantidad de información durante su entrenamiento que ninguna persona podría leer tanto contenido durante su vida. Se basan en probabilidad estadística para la predicción de la siguiente palabra, careciendo de abstracción mental. Por muy humanas que resulten sus creaciones, son incapaces de comprender lo que están expresando^{7,27}, quedando su inteligencia en entredicho. Existen numerosas versiones comerciales al alcance del público, como Claude, DeepSeek, Gemini, GPT, LLaMa, etc.

Visión

Función indispensable para que el cerebro diseñe un modelo subjetivo del mundo que nos rodea, vinculada con la atención y memoria²³. Las aferencias visuales se interpretan a nivel de la corteza occipital, con el apoyo de áreas de localización parietal, temporal y frontal. En mamíferos se ha objetivado que cada neurona o grupo de neuronas reacciona de forma selectiva a estímulos provenientes de regiones concretas del campo visual (campo receptivo)^{19,20,23}. La información lumínica captada por la retina y transformada en impulsos eléctricos alcanza el lóbulo occipital, donde comienza a desgranarse de forma secuencial, extrayendo en primer lugar características simples (orientación espacial, contornos y formas sencillas en áreas V1-V2) y posteriormente las más abstractas (color, movimiento, formas complejas y estereopsis en V3-V6)²³. Las redes neuronales convolucionales (CNN, del inglés *convolutional neural networks*) son arquitecturas de IA que procesan datos de imagen siguiendo una jerarquía similar, destacando en tareas de clasificación y descripción^{19,20}. Al utilizar la representación numérica correspondiente a los píxeles de entrada pueden identificar millones de tonos de gris, a diferencia del ser humano, y aplicar filtros matemáticos que captan los elementos formantes de la imagen (contorno, color, textura...) para obtener un resultado²¹. Son expertas en el análisis de imagen médica, especialmente un subgrupo denominado «redes de cápsulas» (*CapsNets*), introducidas por Geoffrey Hinton, aún en desarrollo²⁰. No obstante, necesitan ser entrenadas con millones de imágenes similares para clasificar de forma adecuada conceptos tan simples como un gato o una

persona⁷. Tanto en seres humanos como en sistemas de IA acontece cierta degradación de la información de imagen durante su transmisión²³.

Aprendizaje y plasticidad neuronal

Aprender involucra numerosos procesos cognitivos (percepción, atención, memoria, toma de decisiones, interpretación de resultados...) y requiere de la integridad del sueño nocturno. Para ello, el cerebro codifica la información recibida de forma subespecializada, habilidad que algunas DNN tratan de replicar^{7,20}. A escala microscópica, el aprendizaje surge de la plasticidad sináptica (modificaciones en los circuitos neuronales), la modulación química mediada por neurotransmisores, la neurogénesis (fundamentalmente en embriones, en adultos solo en localizaciones muy específicas) y la acción de células gliales (liberación/ retirada de neurotransmisores, regeneración de tejido dañado...)²⁸. En el campo de la IA existen infinidad de paradigmas de aprendizaje automático, así que a continuación se menciona una selección de estos.

APRENDIZAJE POR REFUERZO

Se inspira en estudios con modelos animales para alcanzar la máxima efectividad de un modelo al establecer un sistema de recompensa/castigo¹⁹.

APRENDIZAJE SUPERVISADO

El modelo conoce de antemano la respuesta correcta asociada a cada entrada (datos etiquetados). Supone una gran inversión de tiempo y dinero²⁰.

APRENDIZAJE NO SUPERVISADO

No existen datos etiquetados, el modelo solo busca patrones y estructura en los datos (encuentra diferencias). Resulta más eficiente²¹.

ALGORITMO DE RETROPROPAGACIÓN

Permite a las ANN imitar la plasticidad neuronal²⁰. Consiste en un flujo de información en sentido inverso al biológicamente posible (como si el impulso viajara desde el axón hacia las dendritas), que modifica el valor de los pesos sinápticos de la red de forma continua y simultánea en todas las capas, corrigiendo sus errores para acercarse más al resultado final deseado la próxima vez que se enfrente a una tarea^{28,29}. Otro argumento que lo

hace inviable en seres vivos es que la plasticidad depende de eventos locales y la retropropagación lo hace del resultado final de todas las conexiones de la red²⁹; por ello se están estudiando algoritmos de aprendizaje local.

APRENDIZAJE CONTINUO

Implica adquirir nuevas aptitudes sin olvidar las previas. Se postula la existencia de mecanismos especializados a nivel de las espinas dendríticas capaces de proteger determinadas conexiones mientras se establecen otras nuevas. En contexto de DNN se han creado «pesos sinápticos elásticos», que reducen el grado de ajuste de parámetros relevantes para la resolución de problemas aprendidos previamente¹⁹.

REDES GENERATIVAS ANTAGONISTAS

¿Por qué aprender observando si puedes aprender involucrándote activamente en una actividad? Esta es la base del funcionamiento de las redes generativas antagonistas GAN (del inglés *generative adversarial networks*). Consiste en enfrentar una red «aprendiz» a una red «maestro» (que domina la tarea que la otra debe adquirir). Ambas redes reciben *feedback* de forma constante. Con el paso del tiempo, la red aprendiz termina por superar las capacidades de discriminación, producción o juego de su adversaria. En dicho momento se resetea a la red maestro y se invierten los papeles. Al iterar de forma repetida se acaba obteniendo una red altamente especializada con mínima intervención humana²⁷.

REDES NEURONALES DE IMPULSOS

Las redes neuronales de impulsos SNN (del inglés *spiking neural networks*) son un subtipo de DNN fuertemente inspirado en las neuronas biológicas, que incorpora las nociones de tiempo y frecuencia de estimulación. Cada neurona artificial no calcula el sumatorio de sus aferencias continuamente, sino que espera a alcanzar un «potencial de acción virtual» antes de generar una señal de salida, momento en que este se resetea. Especializadas en reconocimiento de patrones dinámicos (voz o imágenes en movimiento), su eficiencia es mayor y acumulan más información que otras DNN. Su elevada complejidad dificulta actualmente el ajuste automático de sus parámetros (se necesitan alternativas a la retropropagación tradicional) y el diseño de modelos a gran escala^{7,20,28}.

Creatividad

En el sentido estricto de la palabra, creatividad es inventiva. Es cierto que las DNN pueden establecer conexiones entre dominios aparentemente no relacionados y desplegar sus habilidades en contextos cambiantes²⁰. Sin embargo, ¿hasta qué punto puede un modelo artificial ser creativo, habiendo completado un entrenamiento exhaustivo con todo el contenido digital disponible? Por otro lado, las emociones tienen un papel crucial en la guía de la expresión creativa, sin nada que apunte a un atisbo de emoción verdadera en el interior de los sistemas de IA actual.

Consumo de energía

Gracias a un largo proceso evolutivo, el cerebro es un órgano extremadamente eficiente (su consumo energético es mínimo en comparación con la complejidad de las operaciones que lleva a cabo)^{20,26}. En este sentido, las ANN tienen mucho que aprender. No es cuestión de cantidad de neuronas, sino de la calidad de su funcionamiento individual y colectivo²².

Patología artificial

Este apartado aborda aspectos potencialmente deficitarios pero inherentes a la propia IA desde una perspectiva metafórica.

Carga cognitiva

Tal y como las personas experimentan sobrecarga cognitiva cuando se enfrentan a un exceso de información simultánea, las ANN pueden alcanzar un límite de procesamiento que les impida completar una tarea correctamente. La escalabilidad es la capacidad de un modelo para procesar datos cada vez más abundantes y resolver problemas más enrevesados²⁰.

Opacidad interna

Las conexiones neuronales sufren constantes modificaciones para optimizar el comportamiento y desempeño humano, siendo todo un enigma el impacto individual de cada uno de esos cambios en la ecuación cerebral global. A su vez, se dice que las DNN se comportan como «cajas negras» debido a la complejidad de su funcionamiento interno, que tampoco ha sido completamente esclarecido. No es posible conocer con exactitud qué ha llevado a un modelo de IA a decidirse por una u otra respuesta, con el potencial riesgo asociado. Inspirándose

en recursos neurocientíficos como el registro unicelular, pruebas de neuroimagen o técnicas lesionales, se pretende desvelar el gran misterio de las ANN^{19,29}.

Fenómeno de alucinación

Sucede cuando un modelo de IA generativa produce información distorsionada (incoherente, incorrecta o inexistente), aparentemente verosímil y no concordante con los datos de su entrenamiento, a pesar de la confianza con la que expresa dicha información. Dada la naturaleza no consciente de estos modelos, las alucinaciones o confabulaciones pueden deberse a una amplia variedad de errores en su programación: sobreajuste, alteraciones lógicas o matemáticas, fallos de razonamiento, etc. Aunque la tendencia parece indicar que los LLM más recientes presentan menor riesgo de cometer alucinaciones, siempre debe contrastarse la información que se obtiene de la interacción con IA, especialmente en el entorno médico³⁰.

Pareidolia

Nuestro cerebro interpreta de forma constante toda clase de información e identifica relaciones relevantes en esta. La pareidolia es la tendencia a percibir patrones familiares en imágenes ambiguas, siendo su variante facial la más frecuente (reconocer caras en imágenes que no contienen rostro alguno). Parece tratarse de una ventaja evolutiva para la supervivencia y la socialización, influida por experiencias vitales y aprendizaje previo, con un papel fundamental en la generación de predicciones. Las áreas temporoparietales e insulares de ambos hemisferios están especialmente implicadas en su aparición, predominando la actividad del lado derecho durante actividades que la evocan³¹. Un estudio con CNN objetivó que aquellas redes entrenadas para identificar tanto rostros como objetos inanimados cometían errores de clasificación al presentarles imágenes que inducen pareidolia en humanos. No obstante, este hecho era mucho menos frecuente entre aquellas redes entrenadas únicamente para discriminar caras³². En contexto de las CNN, la pareidolia puede considerarse fenómeno alucinatorio o recurso artístico generativo. Modelos como DeepDream (Google) se basan explícitamente en este fenómeno para diseñar imágenes a partir de otras^{31,32}.

Selección antinatural

Los novedosos algoritmos genéticos y algoritmos evolutivos se inspiran en el concepto de selección

natural. En el primer caso se compara el rendimiento de diversas ANN en un contexto específico; las más efectivas se seleccionan para «producir descendencia» en una suerte de fusión de código. En el segundo caso se introducen «mutaciones» aleatorias en los pesos sinápticos de varios modelos y se evalúan sus capacidades, escogiéndose solo los mejores. Existe riesgo de inestabilidad o errores inesperados *a posteriori*. En ambos supuestos, las redes artificiales subóptimas son eliminadas²⁸.

Machine un-learning

Del mismo modo que una lesión cerebral induce pérdida de función neurológica, un algoritmo inteligente puede modificarse externamente para alterar su aprendizaje previo, ya se trate de datos específicos contenidos en su memoria o de habilidades de carácter más general. En ocasiones, una lesión estratégica a nivel central tiene carácter terapéutico. Inducir un daño controlado en el código de un modelo de IA puede facilitar la protección de datos de usuarios que no desean contribuir a su entrenamiento²⁰.

Destellos de la singularidad

Hasta ahora se ha abordado el pasado y presente de la IA, por lo que esta sección dirige su enfoque a todo aquello que se atisba en el horizonte tecnológico.

AGI y ASI

El término «inteligencia artificial general» (AGI, del inglés *artificial general intelligence*) hace referencia a un sistema digital capaz de resolver una amplia gama de problemas, incluyendo aquellos que no fueron predeterminados por sus programadores. Exige que el *software* sea capaz de adaptarse al entorno externo, a pesar de partir de una serie de recursos y proceso de entrenamiento limitados, aprendiendo de su experiencia propia y perfeccionando sus capacidades para finalmente tratar de superar a las del ser humano^{6,33}. Sin embargo, esto no implica que deba hacerlo de la forma en que lo haría una persona: aún no se ha establecido si alejarse o adherirse a los de los fundamentos neurocientíficos de la inteligencia humana potenciará o debilitará las capacidades de la IA futura⁶. Cuando la AGI logra exceder las capacidades del ser humano en todos los aspectos de lo que se considera inteligencia, se convierte en súper-IA (ASI, del inglés *artificial superintelligence*), cuyo grado de intelecto hace

hipotéticamente posible la autoconsciencia y percepción de emociones, con el consecuente impacto en el proceso de toma de decisiones^{20,34}. Diversos autores consideran plausible la creación de una AGI, el cuándo y el cómo son la verdadera cuestión³⁵.

Dualismo utopía-distopía

La singularidad es un evento disruptivo de consecuencias irreversibles para el conjunto de la sociedad. En ciencias de la computación, se conoce como singularidad tecnológica al momento en que una IA de carácter general adquiere inteligencia superior a la del humano más inteligente. Una definición alternativa es la pérdida del control del progreso tecnológico al volverse nuestra creación completamente independiente y autorreplicante^{35,36}. La primera persona en utilizar el término en este contexto fue John von Neumann en 1957, aunque no se popularizaría hasta casi 40 años más tarde. Del mismo modo que, hoy en día, los seres humanos no pueden competir de forma exitosa contra ordenadores en la resolución de enrevesados enigmas matemáticos ni a la hora de jugar al póker, al ajedrez o al go, cabe preguntarse si en el futuro la IA no será capaz de sobrepasar todas nuestras capacidades cognitivas. A partir del momento en que la IA supere la capacidad de codificación de los ingenieros que la programaron, la singularidad sería teóricamente posible, trayendo consigo tanto potenciales beneficios (en ámbitos de la salud, educación, finanzas, transporte, gestión de recursos, conflictos globales...) como riesgos muy significativos^{35,37}.

Desafíos pendientes

A pesar de la prometedora trayectoria de la IA, en la actualidad todavía se enfrenta a numerosos obstáculos y arrastra importantes limitaciones que contrastan con la visión idealizada de gran parte del público general.

Entre los aspectos que pulir encontramos la posibilidad de adquirir conocimientos intuitivos, de aprender a partir de ejemplos muy limitados o de generalizar ante situaciones conflictivas no previstas en el entrenamiento, además de la reducción drástica del consumo asociado a la utilización de modelos impulsados por DNN^{19,20}.

En lo referente a riesgos previsibles a corto y largo plazo destacan la magnificación de sesgos de diversa índole introducidos por desarrolladores, filtraciones de datos personales altamente sensibles, desestructuración del mercado laboral, intensificación de las desigualdades económicas, desaparición de la soberanía

humana por delegación en sistemas automatizados (y dilución de responsabilidades asociadas), pérdida de determinadas habilidades instrumentales (lectura, escritura, aprendizaje de idiomas...) o el incremento drástico del consumo de recursos del planeta^{34,37}.

Disponer de marcos legales y éticos adaptativos, así como de un eje de cooperación interdisciplinario neurociencia-ingeniería-gobernanza-ciudadanía que aporte transparencia y establezca un lenguaje común, es un requisito indispensable para asegurar el alineamiento entre intereses humanos y artificiales³⁷.

El futuro es ahora

¿Qué podemos esperar de la IA en los próximos años, en los próximos meses? Si la redacción de este artículo hubiera concluido tan solo una semana antes de lo previsto, esta sección hubiera mostrado un tono diferente, y cuando se publique, un sinnúmero de novedades disruptivas se habrá dado seguramente a conocer. El primer trimestre de 2025 trajo consigo tecnología propia de un largometraje de ciencia ficción, que ve la luz por primera vez tras años de investigación^{6,28}.

COMPUTACIÓN CUÁNTICA

Microsoft presenta Majorana¹, el primer chip cuántico del mundo. Concebido para resolver tareas computacionales extremadamente complejas que supondrían décadas de cálculo a los mejores ordenadores tradicionales disponibles en la actualidad.

COMPUTACIÓN BIOLÓGICA/NEUROMÓRFICA

Cortical Labs anuncia CL1, el primer ordenador biológico comercial, constituido por la fusión de circuitos de silicio y una red neuronal orgánica programable que ha sido cultivada sobre estos. La inteligencia organoide parece ser el paradójico siguiente paso evolutivo de su contrapartida artificial.

CORPOREIDAD DE ARQUITECTURAS DIGITALES

El robot humanoide NEO Gamma, fabricado por 1X, hace gala de sus habilidades de IA para completar tareas domésticas y de acompañamiento de personas.

AGENTES DE IA MULTIMODALES

A diferencia de los modelos de lenguaje tradicionales (carácter informador), Operator de OpenAI es un

software que controla de forma autónoma un buscador online y ejecuta tareas de varias secuencias (comprar un billete de tren, reservar en un restaurante, generar una publicación en red social, etc.) a petición del usuario.

MODELOS CONVERSACIONALES SOFISTICADOS

Con una latencia de respuesta mínima, una gran capacidad de imitación del lenguaje natural y una voz sintética que simula incluso la respiración humana, la demo conversacional de acceso público de la empresa Sesame hace que el test de Turing se tambalee una vez más.

Conclusión: Ctrl+Alt+Supr

A lo largo de las páginas previas se ha expuesto la extensa trayectoria de la IA y su estrecha relación con la neurología, así como el potencial transformador que alberga³⁷. Sin embargo, la búsqueda de AGI parte de motivaciones muy distintas y no todas las predicciones de futuro son precisamente esperanzadoras. Si como sociedad miramos hacia otro lado mientras las nuevas inteligencias y aquellos que creen controlarlas trascienden los límites de la moral y el sentido común, deberemos asumir el riesgo de que la primera entidad digital superinteligente que invente el ser humano sea lo último que cree jamás³⁶. ¿Hemos considerado el grado de dependencia al que nos aboca la IA-cracia? ¿Estamos preparados para dejar de actuar como la medida de todas las cosas?⁴ ¿Sabremos identificar, si fuera necesario, el momento preciso para pulsar el botón de apagado de emergencia?³⁷ ¿De qué sirve preocuparse por el distópico fin de nuestra especie a manos de la IA cuando los humanos somos, potencialmente, nuestro peor enemigo? ¿Existe conciencia del ingente consumo de energía y de recursos naturales que deriva de la utilización irresponsable de las herramientas más populares de IA, tan fácilmente accesibles para el público general?²⁶ A estas y muchas otras cuestiones, solo el tiempo podrá darles respuesta. La IA es neurotecnología en sí misma y el único indicio sólido de la plausibilidad de la AGI/ASI yace escondido en el interior de nuestro cerebro.

Agradecimientos

Al Dr. David Ezpeleta, cuya sesión "Inteligencia Artificial: qué es, qué ofrece y cuáles son sus límites",

impartida en la LXXVI Reunión Anual de la SEN, inspira el hilo conductor del artículo.

Financiación

El presente trabajo no ha recibido ninguna subvención oficial, beca o apoyo de un programa de investigación destinados a la redacción de su contenido.

Conflicto de intereses

El autor no comunica conflicto de intereses en relación con el contenido del trabajo.

Consideraciones éticas

Protección de personas y animales. El autor declara que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad, consentimiento informado y aprobación ética. El estudio no involucra datos personales de pacientes ni requiere aprobación ética. No se aplican las guías SAGER.

Declaración sobre el uso de inteligencia artificial. El autor declara que no utilizó ningún tipo de inteligencia artificial generativa para la redacción de este manuscrito.

Bibliografía

- Grier DA. When computers were human. Princeton, New Jersey: Princeton University Press; 2005.
- Espinosa-Sanchez JM, Gomez-Marin A, de Castro F. The Importance of Cajal's and Lorente de Nó's Neuroscience to the Birth of Cybernetics. *Neuroscientist*. 2025;31(1):14-30.
- Sigman M, Bilinkis S. La génesis de la inteligencia. En: Sigman M, Bilinkis S. *Artificial: la nueva inteligencia y el contorno de lo humano*. Sant Andreu de la Barca, Barcelona: Penguin Random House Grupo Editorial; 2023.
- Lassalle JM. *Civilización artificial*. Barcelona: Arpa Editores; 2024.
- Chen Y, Wei Z, Gou H, Liu H, Gao L, He X, et al. How far is brain-inspired artificial intelligence away from brain? *Front Neurosci*. 2022;16:1096737.
- Xu B. What is meant by AGI? On the definition of Artificial General Intelligence [Internet]. arXiv; 16 abr 2024. Disponible en: <https://arxiv.org/abs/2404.10731>
- Wang G, Bao H, Liu Q, Zhou T, Wu S, Huang T, et al. Brain-inspired artificial intelligence research: A review. *China Technol. Sci*. 2024; 67:2282-96.
- Freeth T, Higgin D, Dacanalis A, MacDonald LW, Georgakopoulou M, Wojcik A. A Model of the Cosmos in the ancient Greek Antikythera Mechanism. *Sci Rep*. 2021;11:5821.
- Xu M, David JM, Kim SH. The Fourth Industrial Revolution: Opportunities and Challenges. *International Journal of Financial Research*. 2018;9:90-5.
- Fidora AM, Sierra C, Barberà S, Beuchot M, Bonet E, Bonner AJ, et al. Ramon Llull: From the Ars Magna to Artificial Intelligence. Barcelona: Instituto de Investigación en Inteligencia Artificial; 2011.
- Grzybowski A, Pawlikowska-Lagód K, Lambert WC. A History of Artificial Intelligence. *Clin Dermatol*. 2024;42(3):221-9.
- Ezpeleta D. Polímatas y diletantes. *Kranion*. 2024;19:41-3.
- Espinosa-Sanchez JM, Perez-Fernandez N, de Castro F, Batuecas-Caletrio A. Cajal's contributions to vestibular research. *Front Neuroanat*. 2024;18:1476640.
- Strawn G. Where Deep Learning and Generative AI Started: Masterminds of Artificial Neural Networks—McCulloch, Pitts, and Rosenblatt. *IT Professional*. 2024;26:99-101.
- Mühlenbein H. Artificial Intelligence and Neural Networks the Legacy of Alan Turing and John von Neumann. *Int J Comput*. 2014;5:10-20.
- Li B, Gilbert S. Artificial Intelligence awarded two Nobel Prizes for innovations that will shape the future of medicine. *NPJ Digit Med*. 2024;7(1):336.
- Cai F, Yuan Y. Who invented backpropagation? Hinton says he didn't, but his work made it popular [Internet]. *SyncedReview*; 24 abr 2020 [citado 9/mar/2025]. Disponible en: <https://medium.com/syncedreview/who-invented-backpropagation-hinton-says-he-didnt-but-his-work-made-it-popular-e0854504d6d1/>
- Yann LeCun: An Early AI Prophet [Internet]. *Dataiku*; 10 abr 2021 [citado 9/mar/2025]. Disponible en: <https://www.historyofdatascience.com/yann-lecun/#:~:text=LeCun%20is%20widely%20credited%20for,able%20to%20identify%20handwritten%20characters/>
- Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-Inspired Artificial Intelligence. *Neuron*. 2017;95(2):245-58.
- Ren J, Xia F. Brain-inspired Artificial Intelligence: a comprehensive review [Internet]. arXiv; 27 ago 2024. Disponible en: <https://arxiv.org/abs/2408.14811>
- Kourounis G, Elmahmudi AA, Thomson B, Hunter J, Ugail H, Wilson C. Computer image analysis with artificial intelligence: a practical introduction to convolutional neural networks for medical professionals. *Postgrad Med J*. 2023;99(1178):1287-94.
- Vishnupriya KP, Jwala J, Prince J, Sriha S, Gibi KS. Brain-Inspired artificial intelligence: revolutionizing computing and cognitive systems. *JSREM*. 2024;08(12):1-6.
- Ye H. Integrating Vision: From Neuroscience to Artificial Intelligence. *Applied and Computational Engineering*. 2024;104:159-64.
- Fabio RA, Capri T, Towey GE, Antonietti A. The neural basis of attention. En: Fabio RA, Capri T, Towey GE, Antonietti A. *Attention today*. New York: Nova Science Publishers Incorporated; 2019.
- Vaswani A, Shazeer NM, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need [Internet]. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. Disponible en: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- Xu B, Poo MM. Large language models and brain-inspired general intelligence. *Natl Sci Rev*. 2023;10(10):nwad267.
- Sigman M, Bilinkis S. Una nueva era. En: *Artificial: la nueva inteligencia y el contorno de lo humano*. Sant Andreu de la Barca, Barcelona: Penguin Random House Grupo Editorial; 2023.
- Schmidgall S, Achterberg J, Miconi T, Kirsch L, Ziaei R, Hajiseyedrazi SP et al. Brain-inspired learning in artificial neural networks: a review. *APL Mach Learn*. 2023;2:2.
- Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G. Backpropagation and the brain. *Nat Rev Neurosci*. 2020;21(6):335-46.
- Sun Y, Sheng D, Zhou Z, Wu Y. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanit Soc Sci Commun*. 2024;11:1278.
- Romagnano V, Kubon J, Sokolov AN, Fallgatter AJ, Braun C, Pavlova MA. Dynamic brain communication underwriting face pareidolia. *Proc Natl Acad Sci U S A*. 2024;121(16):e2401196121.
- Gupta P, Dobs K. Human-like face pareidolia emerges in deep neural networks optimized for face and object recognition. *PLoS Comput Biol*. 2025;21(1):e1012751.
- Bubeck S, Chandrasekaran V, Eldan R, Gehrke JA, Horvitz E, Kamar E, et al. Sparks of Artificial General Intelligence: early experiments with GPT-4 [Internet]. arXiv; 13 abr 2023. Disponible en: <https://arxiv.org/abs/2303.12712>
- Todorova M. Types of artificial intelligence - narrow, general and super AI. Some risks and ethical issues. *Institute for the Study of Society and Knowledge, Ethical Studies*; 2019.
- Lara JA, Lizcano D, Martínez MA, Pazos, J. The advent of technological singularity: a formal metric [Internet]. arXiv; 25 jun 2019. Disponible en: <https://arxiv.org/abs/1907.03841>
- Kendiukhov I. A finite-time technological singularity model with artificial intelligence self-improvement [Internet]. arXiv; 31 ago 2019. Disponible en: <https://arxiv.org/abs/2010.01961>
- Sonko S, Adewusi AO, Obi OC, Onwusinkwue S, Atadoga A. A critical review towards artificial general intelligence: Challenges, ethical considerations, and the path forward. *World J Adv Res.Rev*. 2024;21:1262-8.