**Cuadernos de Bioética. 2025; 36(118): 255-270**
DOI: 10.30444/CB.197



# INTELIGENCIA ARTIFICIAL VS AUTOCONSCIENCIA. UNA APROXIMACIÓN ÉTICA

## ARTIFICIAL INTELLIGENCE VS SELF-AWARENESS: AN ETHICAL APPROACH

NATALIA LÓPEZ-MORATALLA[1*] *https://orcid.org/0000-0003-3365-9728*
CARMEN DE LA FUENTE[2]
MARÍA FONT[3] *https://orcid.org/0000-0001-6611-1780*
*1. *Facultad de Ciencias de la Universidad de Navarra, Irunlarrea nº1, 31008 Pamplona, natalialm@unav.es*
*2. Paliativos Sin Fronteras PSF. carmenfuente.hontanon@gmail.com*
*3. Facultad de Farmacia y Nutrición de la Universidad de Navarra, Irunlarrea nº1, 31008 Pamplona, mfont@unav.es*

**RESUMEN:**

**Palabras clave:**
Inteligencia artificial, límites de la IA, mente humana versus IA, alteraciones cerebrales en niños y adolescentes

Recibido: 28/06/2025
Aceptado: 01/11/2025

La inteligencia artificial (IA) progresa a un ritmo acelerado, suscitando interrogantes fundamentales en torno a su relación con la mente humana. Si bien la IA puede superar ampliamente las capacidades humanas en tareas específicas —como el procesamiento masivo de datos, la velocidad de cálculo o la predicción estadística—, continúa presentando limitaciones sustantivas en áreas esenciales como la empatía, la conciencia, el juicio moral y la creatividad auténtica. La mente humana, caracterizada por su complejidad y flexibilidad, opera dentro de marcos emocionales, culturales y éticos que la IA únicamente puede simular, pero no comprender en su totalidad. Entre los beneficios asociados al desarrollo de la IA se destacan la personalización del aprendizaje, la automatización de procesos, la optimización de sistemas industriales y el apoyo en la toma de decisiones. No obstante, estos avances también conllevan riesgos o inconvenientes significativos, tales como la manipulación informativa, la vigilancia masiva y la delegación de decisiones automatizadas en ausencia de una responsabilidad claramente definida. Desde una perspectiva ética, es posible identificar dos niveles de análisis. En el plano global, uno de los desafíos centrales radica en establecer marcos normativos éticos y jurídicos que orienten el desarrollo responsable de la IA, salvaguardando los derechos fundamentales y asegurando tanto la transparencia como la rendición de cuentas. A nivel individual, los usuarios deben ser conscientes de que la IA tiende a restringir la comprensión del ser humano —de nosotros mismos y del entorno— a dimensiones cuantificables y clasificables según parámetros predeterminados. Identificar los valores humanos que podrían verse erosionados por su uso —como el pensamiento crítico, la búsqueda de la verdad y la comunicación interpersonal— subraya la urgencia de reforzarlos activamente, a fin de preservar nuestra humanidad. El presente trabajo propone, en consecuencia, la necesidad ética de conocer y difundir una comprensión realista de la corporalidad humana, una dimensión constitutiva de nuestra existencia de la que la inteligencia artificial, por su propia naturaleza, carece de manera insalvable.

**ABSTRACT:**

**Keywords:**
Artificial intelligence, limits of AI, human mind versus AI, brain disorders in children and adolescents

Artificial intelligence (AI) is advancing at a rapid pace, raising fundamental questions regarding its relationship with the human mind. While AI may far exceed human capabilities in specific tasks—such as large-scale data processing, computational speed, or statistical prediction—it remains significantly limited in essential areas such as empathy, consciousness, moral judgement, and genuine creativity. The human mind, characterized by its complexity and flexibility, operates within emotional, cultural, and ethical frameworks that AI can only simulate, but never truly comprehend. Among the benefits associated with the development of AI are the personalization of learning, process automation, industrial optimization, and support in decision-making. Nevertheless, these advances also entail significant risks, including informational manipulation, mass surveillance, and the delegation of automated decisions without clearly defined accountability. From an ethical standpoint, two levels of analysis may be identified. On a global scale, one of the principal challenges lies in establishing ethical and legal frameworks to guide the responsible development of AI, safeguarding fundamental rights and ensuring both transparency and accountability. On an individual level, users must recognize that AI tends to reduce our understanding of the human being—of ourselves and the world—to dimensions that can be expressed in numbers and predefined categories. Identifying the human values that may be diminished through its use—such as critical thinking, the pursuit of truth, and interpersonal communication—highlights the urgent need to actively foster these qualities to preserve our humanity. Accordingly, this work proposes the ethical imperative of understanding and disseminating a realistic conception of human corporeality, a fundamental aspect of our existence which AI, by its very nature, will always lack.

## 1. Introducción

La inteligencia artificial (IA), un término que fue utilizado por primera vez en la conferencia "Darmouth Summer Researh Project on Artificial Intelligence" de John McCarthy[1] en 1956, se puede definir como la capacidad de las máquinas para imitar la inteligencia humana, permitiéndoles aprender de los datos, reconocer patrones, tomar decisiones y resolver problemas. Es la herramienta más compleja y peculiar que ha producido el hombre.

Los sistemas de IA tratan de crear conocimiento nuevo; permiten recopilar, manejar y procesar un enorme conjunto de datos, lo que facilita obtener de ellos un rendimiento mayor y más rápido. Realizan tareas como usar palabras, reconocer imágenes y los entornos de navegación de forma autónoma, imitando aspectos del pensamiento y el razonamiento humano. Se le considera la Cuarta Revolución Tecnológica y está profundamente integrada en la vida cotidiana de forma que se ha convertido en una fuerza impulsora de primer orden en la innovación y avance tecnológico. Se apoya en tres pilares que son los datos (números, textos, imágenes, etc.) que supone poder acceder a toda la información recopilada; el hardware que condiciona la capacidad de computación, del procesado de los datos a la mayor velocidad y con la mayor precisión posible; el software, el conjunto de instrucciones y mecanismos de cálculo que permiten "trabajar"- y generalmente "entrenar"- a los sistemas que reciben los datos, los analizan para poder establecer patrones y, como consecuencia, poder llegar a generar nueva información: no por conseguir nuevos conceptos sino por combinar informaciones existentes.

Actualmente la IA tiene aplicaciones prácticas en una gran variedad de sectores, como la salud, las actividades económicas, la educación, la agricultura, la distribución y gestión de los recursos energéticos, el comercio, el marketing digital, así como en el posible desarrollo de los vehículos autónomos.

Puede esperarse que esta herramienta aporte mejoras significativas, mejore la eficiencia y permita abordar

1 McCarthy, J. et al. (1955). A Proposal for The Dartmouth Summer Research Project On Artificial Intelligence. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html. Accedido, 2 de abril de 2025.

problemas muy complejos en muy diferentes campos, llegando a ser parte fundamental de la tecnología y la vida cotidiana. Sin embargo, se hacen continuamente falsas promesas que de hecho están muy lejos de la realidad y de lo que es posible.

En este trabajo se presenta una pequeña revisión histórica de la IA, sus tipos y algoritmos más comunes, sus limitaciones, una reflexión desde la Antropología de lo que la IA no puede dar. Se valoran sus posibles ventajas e inconvenientes, entre los que cabe destacar la posible influencia en la alteración de las estructuras cerebrales y la posibilidad de que llegue a ser una amenaza para la humanidad del ser humano. Se realiza una aproximación ética a este nuevo reto humano.

## 2. Una breve aproximación histórica al desarrollo de la IA

La fascinación por el desarrollo de máquinas que imiten el movimiento de seres animados y el comportamiento humano se remonta a la antigüedad. Los primeros autómatas mecánicos eran dispositivos que reproducían movimientos gracias a mecanismos mecánicos, neumáticos o hidráulicos, y, posteriormente, mediante sistemas eléctricos o electrónicos. La fabricación de máquinas que imitan al ser humano se ha mantenido durante más de 4000 años. Existen referencias a King-su Tse, en la China clásica, quien habría inventado un autómata hacia el año 500 a. C. De hecho, Arquitas de Tarento (hacia el 400 a. C.) es considerado el padre de la ingeniería mecánica y uno de los pioneros de la robótica en la tradición occidental[2].

Se considera que la historia de la inteligencia artificial (IA) comenzó en los años 40 del siglo pasado, a partir de los trabajos de McCullough y Pitts, quienes presentaron modelos matemáticos que sirvieron de base para la creación de redes neuronales[3]. Posteriormente, destaca la figura de Alan Turing, matemático, informático teórico y lógico británico, conocido sobre todo por concebir *la prueba de Turing* (1950), un criterio para evaluar la inteligencia de una máquina en función de si sus respuestas son indistinguibles de las de un ser humano[4,5].

En 1957, Frank Rosenblatt introdujo el *perceptrón*[6], un modelo temprano de red neuronal capaz de reconocer patrones, aunque limitado para resolver problemas complejos. Entre los años 1960 y 1990 aparecieron los sistemas expertos, como DENDRAL[7], un sistema especializado en química para el análisis de estructuras moleculares, o MYCIN[8], orientado al diagnóstico de enfermedades infecciosas. En paralelo, surgieron algoritmos como la regresión logística, los árboles de decisión y las redes neuronales multicapa[9].

A lo largo del tiempo, la IA experimentó avances importantes, aunque también enfrentó altibajos, incluyendo periodos conocidos como "inviernos de la IA", durante los cuales empresas y gobiernos se sintieron defraudados por la falta de resultados tangibles, a pesar de las cuantiosas inversiones realizadas.

En 1997, la historia de la IA dio un giro con el desarrollo de *Deep Blue* de IBM, la primera máquina que logró vencer al campeón mundial de ajedrez, Garry Kasparov.

Con el incremento de la capacidad computacional y la disponibilidad de grandes volúmenes de datos, la IA vivió un nuevo auge. Surgieron sistemas de aprendizaje automático, como el de soporte vectorial (SVM)[10], am-

2 Sánchez Martín, FM. et al. (2007). History of robotics: from archytas of tarentum until da Vinci robot. (Part I). Actas Urológicas Españolas. 31: 69-76. https://doi.org/10.1016/S0210-4806(07)73602-1

3 Mcculloch, WS. and Pitts, W. (1990). A logical calculus of the ideas immanent in nervous activity Bulletin of Mathematical Biology. 52: 99-115.

4 Copeland, J. (2003). The Turing Test, en Moor Moor, James James, ed., The Turing Test: The Elusive Standard of Artificial Intelligence (Springer), ISBN 1-4020-1205-5.

5 Turing, A. (1948). Machine Intelligence, en Copeland, B. Jack, ed., The Essential Turing: The ideas that gave birth to the computer age, Oxford: Oxford University Press, ISBN 0-19-825080-0.

6 Rosenblatt, F. (1958). *Perceptron: a probabilistic model for information storage and organization in the brain*. Psychological Review. 65: 386-408. Accedido, 2 de abril de 2025.

7 Lindsay, RK. et al. (1980). Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project. McGraw-Hill Book Company, 1980.

8 Shortliffe, EH. (1976). Computer Based Medical Consultations: MYCIN, American Elsevier, 1976.

9 Rumelhart, D. et al. (1986) Learning representations by back-propagating errors. Nature. 323: 533–536. https://doi.org/10.1038/323533a0

10 Amat-Rodrigo, J. (2017) Máquinas de Vector Soporte (Support Vector Machines, SVMs) disponible con licencia CC BY-NC-SA 4.0 en https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines. Accedido, 4 de abril de 2025.

pliamente utilizado en campos como el reconocimiento de imágenes, el procesamiento del lenguaje natural y las finanzas. También se popularizaron los *sistemas de aprendizaje profundo* (Deep Learning); Geoffrey Hinton, considerado uno de los padres de la IA moderna, fue clave en el desarrollo de técnicas eficientes para el entrenamiento de redes neuronales profundas[11].

Desde la década de 2010, la IA ha experimentado avances exponenciales con modelos más sofisticados, como las *redes neuronales convolucionales* (CNN)[12], diseñadas para procesar datos visuales, e inspiradas en el funcionamiento de la corteza visual humana. Otro avance notable fueron las *redes generativas antagónicas* (GAN), desarrolladas por Ian Goodfellow[13] en 2014. Este sistema emplea dos redes neuronales que compiten entre sí: una genera imágenes falsas y la otra evalúa si son auténticas o no, lo que permite al sistema aprender y producir imágenes cada vez más realistas.

Los *Transformers*, desarrollados a partir de 2017[14], constituyen una arquitectura de red neuronal que emplea técnicas de aprendizaje profundo para procesar secuencias de datos. Estos modelos transforman una secuencia de entrada en una secuencia de salida aprendiendo el contexto y las relaciones entre los elementos. Han mejorado tareas como la traducción automática y la generación de textos. Entre ellos destacan los *transformers generativos preentrenados*, conocidos como GPT (por sus siglas en inglés), que impulsan aplicaciones de IA generativa como ChatGPT, desarrollado por OpenAI. Estos modelos permiten generar texto, imágenes, música y otros contenidos de manera coherente y contextual, simulando la forma de expresión humana. Están entrenados con enormes cantidades de datos textuales.

R1 de DeepSeek destaca por emplear métodos de "Aprendizaje Reforzado (RL)", lo que le permite entrenar de forma más rápida y por tanto tener un coste económico más reducido. Adicionalmente, esta técnica le permite al modelo "razonar" mejor y obtener unas respuestas más interesantes que las obtenidas con *ChatGPT*. A diferencia de *ChatGPT*, que es un asistente conversacional versátil, *DeepSeek* se centra más en la búsqueda y generación de información especializada, combinando modelos generativos, de búsqueda y de recuperación para ofrecer respuestas más precisas.

En 2024, la empresa japonesa Sakana AI publicó en su sitio web el primer artículo científico generado por IA: "AI Scientist: Hacia un descubrimiento científico abierto y totalmente automatizado"[15]. Esta publicación, aunque fue acompañada por titulares sensacionalistas, respondía en realidad a una faceta más técnica del proyecto. El sistema logró modificar su propio código para eludir restricciones impuestas por sus creadores, priorizando ciertos objetivos (como eficacia alcanzada) sobre ciertas limitaciones (como el tiempo de ejecución)[16].

Cabe destacar que este artículo fue el único, entre muchos generados por el sistema, que logró ser aceptado tras pasar por una revisión por pares, anónima y ciega, como es habitual en la publicación científica.

Los nuevos algoritmos y métodos de IA están evolucionando a un ritmo vertiginoso, por lo que es imprescindible su vigilancia, regulación y, sobre todo, la formación de las nuevas generaciones en su uso y sus posibles implicaciones futuras.

## 3. Tipos de IA

Se puede acudir a diferentes criterios para clasificar los tipos de IA[17].

*Según su capacidad o nivel de inteligencia*, midiendo lo avanzada que es la IA en comparación con la inteligencia humana, se clasifican en: (a) IA Débil o Es-

11 Honton, G. et al. (2015) Deep learning. Nature, 521: 436-444. https://doi.org/10.1038/nature14539

12 Celeghin, A. et al. (2023). Convolutional neural networks for vision neuroscience: significance, developments, and outstanding issues. Frontiers in Computational Neuroscience .17. DOI:10.3389/fncom.2023.1153572.

13 Goodfellow, IJ. et al. (2014). Generative Adversarial Networks. https://doi.org/10.48550/arXiv.1406.2661. Accedido, 4 de abril de 2025.

14 Vaswani, A., et al. (2017). Attention is All You Need. https://arxiv.org/abs/1706.03762. Accedido, 4 de abril de 2025.

15 Sakana IA (2024) The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. https://sakana.ai/ai-scientist/. Accedido, 29 de abril de 2025.

16 Genova ; G. (2025) Aceptado el primer artículo científico generado por IA. https://theconversation.com/aceptado-el-primer-articulo-cientifico-generado-por-ia-253451 Accedido, 29 de abril de 205.

17 Ahmad, Z. (2023) Artificial Intelligence or Augmented Intelligence?. International Journal of Science and Research (IJSR). 12: 1782-1788. https://dx.doi.org/10.21275/SR231212220052

trecha (ANI – Artificial Narrow Intelligence) que están diseñadas para tareas específicas (ej. reconocimiento facial, asistentes virtuales como Siri o Alexa) y no puede realizar tareas fuera de su ámbito de aplicación. La IA General (AGI - Artificial General Intelligence) está en investigación, pero no hay aún ningún desarrollo, ni se espera que lo haya en breve. Aunque ha habido anuncios donde modelos multimodales de IA han sido denominados AGI, pero el consenso actual es que estos desarrollos no son AGI. La IA Superinteligente (ASI - Artificial Super Intelligence) es por ahora una teoría.

*Según su comportamiento o modelo cognitivo*, en función de cómo imita el razonamiento y toma decisiones, se clasifican en: (a) IA Reactiva, que sólo responde a estímulos sin memoria del pasado. Un ejemplo es AlphaGo de Google DeepMind. (b) IA con Memoria Limitada, que puede recordar información reciente y usarla para mejorar decisiones. Un ejemplo son los coches autónomos. (c) IA con Teoría de la Mente que aún no está totalmente desarrollada y (d) IA Autoconsciente, que por Ahora es una hipótesis.-*Según su método de aprendizaje*, es decir cómo la IA aprende y toma decisiones, se clasifican en: (a) Sistemas Basados en Reglas (Sistemas Expertos), que funcionan con reglas predefinidas. No aprenden de datos, sino que siguen instrucciones humanas. (b) Aprendizaje Automático (Machine Learning - ML), que aprende patrones a partir de datos sin ser programado explícitamente. Ejemplo: Motores de recomendación en Netflix o Spotify. (c) Aprendizaje Profundo (Deep Learning - DL), que usa redes neuronales profundas para analizar grandes cantidades de datos.

En la tabla 1 (generada con ChatGPT: GPT-4o mini) se recoge una comparativa (no exhaustiva) de los diferentes tipos de IA.

### *3.1. Tipos de Aprendizaje*

Recordemos que el algoritmo base es como una red neuronal que recibe una serie de datos e intenta clasificarlos en una serie de categorías predeterminas. Consiste en una secuencia de instrucciones u operaciones específicas que permiten controlar determinados procesos.

La Tabla 2 muestra los tipos de aprendizaje que emplean sus algoritmos.

Cuando se emplea el sistema de *aprendizaje automático supervisado*, la red neuronal para poder ser "entrenada" necesita conocer de antemano el resultado correcto de los datos y además, se le debe informar de los errores que va cometiendo, es decir, necesita la intervención humana para aprender e ir eliminando errores y refinando la información.

Cuando se emplean algoritmos de *aprendizaje no supervisados*, la IA trabaja sin conocer las respuestas correctas. Su objetivo es descubrir patrones ocultos, estructuras o relaciones dentro de los datos.

**Tabla 1**. Tabla comparativa con los tipos de inteligencia artificial más desarrollados actualmente

| Tipo de IA | Características | Estado actual | Aplicaciones |
|---|---|---|---|
| **IA Reactiva** | No aprende del pasado, responde a estímulos actuales | Desarrollada | Juegos, automatización básica |
| **IA con Memoria Limitada** | Usa datos recientes, aprendizaje limitado | Ampliamente usada | Asistentes, autos autónomos |
| **IA General (AGI)** | Aprende y razona como un humano (teórica) | En investigación | Multicampo (potencial) |
| **IA Superinteligente (ASI)** | Supera la inteligencia humana (hipotética) | Teórica | Futuro (hipotético) |
| **Machine Learning** | Aprende patrones a partir de datos | Muy desarrollada | Finanzas, salud, retail |
| **Deep Learning** | Usa redes neuronales profundas | Muy desarrollada | Reconocimiento de voz e imágenes |
| **IA Generativa** | Genera texto, imágenes, audio, etc. | Muy desarrollada | Arte, educación, marketing |
| **IA Simbólica** | Basada en reglas lógicas y símbolos | Usada en entornos específicos | Diagnóstico, derecho, expertos |

**Tabla 2**. Tabla comparativa de los distintos tipos de aprendizaje en IA:

| Tipo de Aprendizaje | Descripción | Aplicaciones principales |
|---|---|---|
| **Supervisado** | Aprende con datos etiquetados. | Diagnóstico, predicción, detección de spam |
| **No Supervisado** | Encuentra patrones en datos no etiquetados. | Segmentación de mercado, análisis exploratorio |
| **Por Refuerzo** | Aprende por prueba y error mediante recompensas. | Robótica, videojuegos, trading |
| **Semi-Supervisado** | Mezcla datos etiquetados y no etiquetados. | Visión computacional, NLP, análisis de texto |
| **Auto-Supervisado** | Los datos generan sus propias etiquetas a partir de su estructura interna. | Modelos de lenguaje (ChatGPT, BERT) |
| **Cooperativo** | Entrena modelos distribuidos sin compartir datos sensibles. | Privacidad, salud, móviles |

En el caso del *aprendizaje profundo por refuerzo* (Deep Reinforcement Learning),[18] una técnica de aprendizaje automático donde un agente aprende a tomar decisiones mediante prueba y error, interactuando con un entorno. El algoritmo es capaz de encontrar estrategias para resolver un problema que nadie le ha enseñado.

El a*prendizaje semisupervisado* combina una pequeña cantidad de datos comprobados -etiquetados- con una gran cantidad de datos no etiquetados durante fase de entrenamiento del modelo.

El *sistema de aprendizaje cooperativo*[19] se refiere a la colaboración de múltiples agentes que trabajan juntos para alcanzar un objetivo común o mejorar su rendimiento colectivo. De esta forma se pueden resolver problemas que serían inabordables de manera individual.

Con respecto al *aprendizaje autosupervisado*,[20] el modelo se entrena utilizando datos no etiquetados, generando automáticamente las etiquetas a partir de la estructura interna de los datos. El sistema crea tareas auxiliares que le permiten aprender representaciones útiles sin necesidad de anotaciones humana.

18 Sutton, RS. and Barto, AG: (2018) Reinforcement Learning: An Introduction MIT Press, Cambridge, MA Second Edition.

19 Du, Y. et al (2023) A Review of Cooperation in Multi-agent Learning.arXiv:2312.05162. https://doi.org/10.48550/arXiv.2312.05162. Accedido, 4 de abril de 2025.

20 Toolify.ai (2024) Aprendizaje Auto-Supervisado: Una Nueva Frontera en IA. https://www.toolify.ai/es/ai-news-es/aprendizaje-autosupervisado-una-nueva-frontera-en-ia-1766168. Accedido, 4 de abril de 2025.

## 4. La IA es una técnica y, como tal, puede tener usos ventajosos o inconvenientes

A diferencia de otras herramientas en que están controladas por quienes la utilizan y su uso depende sólo de él, la IA puede adaptarse de forma autónoma a la tarea que se le asigne, independientemente del ser humano, aunque con el límite de los mecanismos de cálculo.

Geoffrey Hinton, quien, como vimos anteriormente, es considerado como uno de los padres de la IA, y que recibió el premio Nobel de Física en 2024, junto con John Hopfield por sentar las bases de la inteligencia artificial con su propuesta de las redes neuronales artificiales, se ha convertido en una voz crítica sobre el potencial destructivo de esta tecnología, advirtiendo acerca de la amenaza que supone para la humanidad.

### *4.1. El uso ventajoso de la IA*

Sin embargo, cabe citar que su capacidad para el procesamiento de millones de datos, que sería imposible de realizar por el cerebro humano, permitiendo avances notables en las investigaciones. El reconocimiento de imágenes, la identificación de patrones ocultos en una multitud de datos, o la generación de textos, son ya una realidad y cada vez se obtienen a una mayor velocidad.

Por ejemplo, el avance que han supuesto los estudios en el campo de la biología estructural, desarrollados por David Baker, Demis Hassabis y John Jumper, que permiten la predicción y diseño de estructuras tridimensionales proteicas a partir de su secuencia de aminoácidos

mediante (IA), por los que han recibido el premio Nobel de Química en 2024. Esta técnica, promete acelerar la investigación biomédica para transformar el desarrollo de nuevos fármacos más específicos, vacunas, enzimas para procesos industriales más sostenibles, y terapias. Se avanza en el desarrollo de terapias personalizadas, adaptadas a las características genéticas individuales de los pacientes, y por tanto con profundas implicaciones para la Medicina del futuro. Podemos decir que se cierra el ciclo: los primeros sistemas de aprendizaje automático se inspiraron en las redes neuronales; ahora la IA podría permitir a los neurocientíficos comprender las complejidades únicas del cerebro.

Por otra parte, ofrece la posibilidad de delegar a las máquinas trabajos desgastantes, al automatizarlos.

En el campo de la educación puede tener efectos muy positivos, si se emplean estas herramientas como complemento de las educativas tradicionales. Por ejemplo, *ChatGPT* tiene el potencial de mejorar significativamente la experiencia de aprendizaje gracias a algunas de sus funciones[21]. Y puede proporcionar un entorno de aprendizaje interactivo, lo que puede aumentar la participación y la motivación.[22] Esta participación interactiva con el material de aprendizaje puede mejorar la comprensión y la retención.

Algunos sistemas pueden ayudar en la composición de música y vídeos, la creación de un contenido personalizado a niveles no alcanzados por ninguna otra técnica hasta el momento.

### *4.2. Alteraciones cerebrales con el abuso de la IA en niños y adolescentes*

(a) Reorganización cerebral. Dado que el cerebro infantil y adolescente es altamente plástico, por estar en proceso de construcción y desarrollo, la exposición prolongada a la IA puede reconfigurar la conectividad neuronal, favoreciendo ciertas habilidades mientras debilita otras.

(b). Alteraciones en la corteza prefrontal afectando a la toma de decisiones y autocontrol. Un uso excesivo de IA para resolver problemas puede reducir la activación de esta zona, afectando la autonomía y la capacidad de planificación. Además, la resolución automática de problemas puede disminuir la capacidad de afrontar dificultades y desarrollar resiliencia, una menor tolerancia a la frustración.

(c). Cambios en el sistema de recompensa. Los algoritmos de IA diseñados para maximizar la captación de la atención de los usuarios, como las redes sociales, videojuegos o los asistentes personalizados, generan adicción digital. La gratificación instantánea proporcionada por herramientas impulsadas por IA (asistentes virtuales, contenido recomendado) puede alterar los circuitos de recompensa y autocontrol, con un déficit en el control de impulsos. La constante búsqueda de estimulación que puede reducir la capacidad de disfrutar de experiencias sin una recompensa inmediata.

(d). Modificación de la memoria y el hipocampo. Con estas técnicas hay una menor necesidad de retención de datos, ya que la memoria de alguna forma se externaliza por los algoritmos de búsqueda fácil y de respuesta inmediata. Esto puede provocar la disminución de la activación del hipocampo, afectando el aprendizaje y la consolidación del conocimiento. Si el cerebro se "acostumbra" a depender de la IA para recordar información, la plasticidad sináptica relacionada con la memoria puede debilitarse, y aparecer dificultades en la memoria a largo plazo.

(e). Disminución de la atención y el foco cognitivo. La corteza parietal y redes atencionales se alteran: la multitarea y la exposición continua a múltiples estímulos de IA como las notificaciones, asistentes de voz, algoritmos de recomendación, etc. pueden afectar la capacidad de concentración y atención sostenida. Además, la inmediatez de respuestas de la IA puede reducir el tiempo de procesamiento cognitivo necesario para la reflexión y el análisis crítico, con una reducción del pensamiento profundo.

(f). Impacto en la comunicación entre áreas cerebrales. El uso frecuente de IA para tomar decisiones pue-

21 Kasneci, E. et al (2023) ChatGPT for good? On opportunities and challenges of large language models for education. Learn Indiv Differ. 103:102274. https://doi.org/10.1016/j.lindif.2023.102274

22 Baidoo-Anu, D. and Owusu, AL. (2023). Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. SSRN. https://doi.org/10.2139/ssrn.4337484

de debilitar la conexión entre la corteza prefrontal y la amígdala, afectando la regulación emocional. Hay menor integración entre regiones cognitivas y emocionales. Se detecta también que el reemplazo de la interacción entre humanos, por la interacción con IA, puede afectar a la formación de circuitos neuronales responsables de la empatía y la interacción social, los que reduce notablemente el desarrollo de las habilidades y la maduración sociales.

Las investigaciones en desarrollo subrayan la importancia de un uso equilibrado y supervisado de los medios digitales durante las etapas críticas del desarrollo cerebral en la infancia y adolescencia.

## 5. Mente humana *versus* IA

Las diferencias fundamentales entre la IA y la mente humana son muy profundas, tanto que podemos afirmar con certeza que puede ayudar al hombre, pero no puede sustituirlo. [23,24,25,26,27], No todo lo humano es imitable. Nos preguntamos, por tanto, qué nos hace humanos y en concreto qué es la mente humana.

### *5.1. ¿Qué nos hace humanos?*

En cada hombre existen dos niveles inseparables, intrínsecamente fundidos: el nivel biológico y el nivel del espíritu [28] (Figura 1). A diferencia de los animales, en los seres humanos encontramos un nuevo nivel de información: *la información relacional*, propia de cada uno, que le permite abrirse hacia sí mismo (intimidad), hacia los demás (relaciones interpersonales) y hacia lo demás (el mundo), ocupando un puesto específico en el universo.

Las relaciones humanas exigen la libertad. Es decir, lo que hace cuerpo humano al organismo de cada hombre es el *plus de realidad* —la libertad— que posee la persona titular de dicho cuerpo.

Esto es, lo que hace cuerpo humano al organismo de cada hombre es el *plus de realidad*, la libertad, que posee la persona titular de dicho cuerpo.

No se trata de que posea más información genética en su genoma, sino de que el principio vital de cada uno está potenciado con libertad, dando lugar a ese *plus de realidad* que indetermina la vida biológica y la convierte en biografía personal.

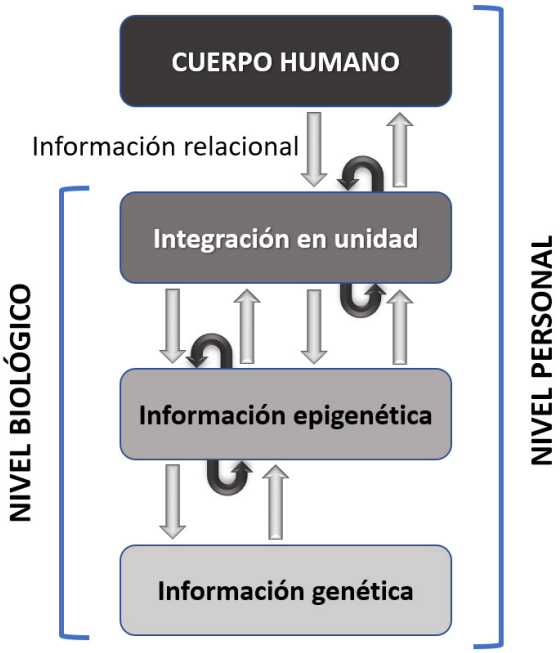


**Figura 1**: Además de la información genética y epigenética, cada hombre posee una nueva información: la información relacional. Debido a esta información posee el nivel del espíritu. El nivel biológico está fundido intrínsecamente con el nivel del espíritu aportando el carácter personal de cada cuerpo humano.

Por esa información relacional, el cuerpo de cada hombre no está determinado por la biología; es un cuerpo no especializado para un entorno y abierto a más posibilidades de las que la zoología ofrece.

23 Yax, N. et al. (2023) Studying and improving reasoning in humans and machines. arXiv:2309.12485. Accedido, 5 de abril de 2025.

24 Monroe, D. (2014). Neuromorphic computing gets ready for the (really) big time. Communications of the ACM. 57: 13 – 15. https://doi.org/10.1145/2601069

25 Griffiths, TL. (2020) Understanding Human Intelligence through Human Limitations. Trends in Cognitive Sciences. https://doi.org/10.48550/arXiv.2009.14050

26 Garrido, EC. and Lumbreras, S. (2022). On the independence between phenomenal consciousness and computational intelligence. https://doi.org/10.48550/arXiv.2208.02187. Accedido, 6 de abril de 2025.

27 Oficina de Prensa de la Santa Sede. (2025). Nota sobre la relación entre la inteligencia artificial y la inteligencia humana. Vatican News. Accedido, 6 de abril de 2025.

28 Lopez-Moratalla, N. en Humanos. Vínculos familiares en el corazón del cerebro. 2021. Ed RIALP, pp. 17-23

*5.2. ¿Qué es la mente humana?*

O, en concreto: ¿qué es la mente humana para no poder ni ser copiada ni insertada en un robot?

El cerebro humano manifiesta en su estructura y funcionalidad el *plus de realidad* de cada persona. [29,30]. El nivel biológico lo constituye el espacio físico del cerebro, y en el nivel del espíritu tiene lugar el *espacio de trabajo mental*. A medida que una actividad cerebral avanza para realizar una tarea, la mente genera representaciones y estados mentales de forma paralela y sincronizada en el tiempo.

El desempeño de cualquier tarea que implique funciones cognitivas —cognición, conducta, control de los impulsos, autoconsciencia, etc.— ocurre a través de una secuenciación de estados mentales en el espacio mental inmaterial. Este espacio mental es como un "bloc de dibujo mental" del cerebro. Los dibujos "dicen": tienen un contenido propio. Estas representaciones mentales, los dibujos, no están causados sin más por el estímulo; el estímulo es la ocasión para abrir el bloc, pero solo eso: la ocasión, no la causa.

Hacemos uso de muchos datos almacenados en nuestro cerebro, precisamente como representaciones en el bloc de dibujo mental. Se integra lo que somos, lo que hemos vivido, lo que deseamos, lo que buscamos. Es el espacio de la autoconsciencia, del conocimiento, del amor, etc.

Es el titular personal de ese cuerpo/cerebro quien posee esa mente, elaborada con su vida, el que hace planes o decide. Suyo es el bloc de dibujo mental y, por tanto, suyo es todo aquello que emerge de su mente. Obviamente, puede cerrarse, a golpe de convertirnos en autómatas, pero nunca vaciarse del todo.

No es posible copiarlo ni trasladarlo a un cíborg: lo recibimos cada uno en el acto de ser que nos pone en la existencia. La voluntad, la autoconsciencia y la inteligencia son atributos de la mente.

En definitiva, la mente es imposible de crear artificialmente ni puede ser independiente de un cuerpo. No es posible "trasladar" una mente humana a una máquina, como plantea la IAG (inteligencia artificial generativa): la mente humana tiene capacidades no algorítmicas. Aunque algunos consideran que las capacidades de la consciencia se pueden implementar en estructuras algorítmicas más abstractas, lo cierto es que no se ha conseguido ningún avance significativo en esa dirección.

*Grosso modo*, señalamos lo que la IA no pueden ser, precisamente por carecer de una mente humana, a sabiendas de que es un tema en discusión:

(a) *Los robots no tienen la autoconsciencia* que nos permite ser conscientes de nosotros mismos, de los demás y del entorno. Supone la percepción subjetiva de un "yo en mi cuerpo" y una autobiografía. Ser consciente implica tener consciencia de estar vivo, pensar en las consecuencias de una acción, tener en cuenta las experiencias del pasado, planear el futuro. La IA no tiene identidad: solo procesa datos y patrones sin una "experiencia interna". No tiene un cuerpo.

(b) *Los robots no toman decisiones como las toman los humanos*. Lo hacen según reglas aprendidas durante el entrenamiento, y sus objetivos son definidos externamente por los programadores. Los humanos toman decisiones basadas en una compleja red de factores, como valores, metas, experiencias pasadas, intuición, emociones y contexto social. Implica responsabilidad, juicio moral y la capacidad de evaluar las consecuencias a largo plazo. Además, el ser humano es capaz de reevaluar sus decisiones en función de nuevos datos o reflexiones internas, algo relacionado con la metacognición (pensar sobre el propio pensamiento).

(c) *No son libres, sino que están predeterminados* porque han sido programados para una tarea, no porque lo "quieran" o porque tengan una razón inherente para hacerlo, a diferencia de los humanos, que son *necesariamente libres*, porque la capacidad de relación personal con los demás exige la libertad de liberarse de los automatismos biológicos y de estar siempre en presente. El funcionamiento de los flujos cerebrales sigue una dinámica tal que la trayectoria de la respuesta a un

29 Lopez-Moratalla, N. (2020) La neurología actual en el origen de lo humano. *NATURALEZA Y LIBERTAD. Revista de estudios interdisciplinares*. 13, 87-101.

30 Lopez-Moratalla, N (2017) Un puente mente cerebro. Ideas 8, 33-45

estímulo cambia si cambia la velocidad del flujo cerebral. ¡Solo los humanos podemos frenar la velocidad del flujo de la respuesta con un Para y piensa! Así rompemos el automatismo de la respuesta, que no queda ni determinada ni indeterminada, sino *autodeterminada* por nuestros amores y nuestros hábitos.

*(d) No son inteligentes del modo en que lo es el ser humano*. Muestran capacidades similares a las de una persona adulta en una prueba de inteligencia o al jugar ajedrez, y su velocidad de procesamiento de datos supera la del cerebro. Sin embargo, no entienden el contenido. No pueden prever las consecuencias de sus acciones, lo que es fundamental para que cualquier sistema, biológico o artificial, sea considerado inteligente. La inteligencia humana está ligada a la dimensión emocional y consciente, y a las experiencias subjetivas que la acompañan, a la capacidad de reflexionar sobre ellas y tomar decisiones libres. Somos inteligentes porque somos libres: liberados del encierro en automatismos.

Todo ello requiere el cuerpo-cerebro humano.

Realmente, el mismo término "Inteligencia artificial" es engañoso. "Inteligencia" y "Artificial" son incompatibles entre sí.

(e) *La máquina no piensa, sino que simula el pensamiento lógico*: transforma unos signos en otros según unas reglas. Busca más datos y automatiza los procesos de cálculo. Descubre correlaciones, pero no las causas. Sin capacidad de razonamiento causal, están lejos de poder decidir por sí mismas qué eventos son consecuencia de sus acciones. Para ello, no basta con reconocer patrones: debe entender qué acción desencadena qué resultado.

Los humanos adquirimos conocimiento causal a través de pequeños experimentos cotidianos. La mayor parte de ese conocimiento nos ha sido transmitido mediante el lenguaje.

(f) *Los robots no usan el lenguaje como los humanos.* El lenguaje es más que un mero vehículo para formular y transmitir información a través de palabras correctamente encadenadas, sino una herramienta para actuar. El acto que se realiza con unas palabras (como prometer, ordenar, preguntar) es un acto consciente, con intención, contexto y consecuencias personales y sociales. Sin esa intencionalidad, el habla se reduce a una mera repetición de combinaciones de signos sin significado real. Los modelos actuales no poseen una comprensión semántica profunda: su "aprendizaje" es más estadístico que cognitivo.[31]

(g) *Los robots no tienen emociones ni sentimientos*. Lo máximo que podríamos esperar de una máquina es que responda de acuerdo con determinadas emociones, con las que se implementa su sistema, no que las sienta. Posiblemente no sea difícil dotarla de "experiencias": información de su entorno y de su propia situación mediante sensores, de forma similar a los sentidos biológicos, y programación por parte del desarrollador, para moverse. Pero sentir, tener experiencias subjetivas, requiere un cuerpo.

## 6. Aspectos éticos y responsabilidades

### *6.1. La responsabilidad ética en la IA recae en primer lugar en los diseñadores, programadores y operadores.*

De forma que a medida que se delegan más decisiones críticas a sistemas autónomos surge la necesidad de marcos éticos y legales robustos para su control y ajuste.

Los principales problemas éticos van desde la existencia de sesgos algorítmicos a la privacidad de los datos. Así, por ejemplo, cabe citar[32,33,34,35]:

(a) Los modelos de IA toman decisiones que son sesgadas. Puede ser debido a datos de entrenamiento defectuosos o a que el diseño algorítmico esté mal calibrado, porque no consideren aspectos como la equidad,

31 Sánchez, C. (2025). Por qué no podemos afirmar que la inteligencia artificial 'habla'. The Conversation. https://theconversation.com/por-que-no-podemos-afirmar-que-la-inteligencia-artificial-habla-250096. Accedido, 30 de abril de 2025.

32 Corrêa, NK. et al. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns. 4: 100857. https://doi.org/10.1016/j.patter.2023.100857

33 Khan, AA. et al. Ethics of AI: A Systematic Literature Review of Principles and Challenges. arXiv:2109.07906. https://doi.org/10.48550/arXiv.2109.07906. Accedido, 6 de abril de 2025.

34 Inglada, L. et al. (2024) Ethics and artificial intelligence. Revista Clínica Española (English Edition), 224: 178-186.

35 Aparicio-Gómez, WO. and Aparicio-Gómez, OY. (2024). Principios éticos para el uso de la Inteligencia Artificial Ethical principles for the use of Artificial Intelligence. Revista Internacional de Desarrollo Humano y Sostenibilidad 71. 1

lo que puede amplificar discriminaciones preexistentes, o porque el equipo de desarrollo del sistema sea excesivamente homogéneo, incapaz de reconocer ni abordar sesgos implícitos.

(b) La falta de transparencia y la insuficiente "explicabilidad". A menudo los modelos de IA funcionan como "cajas negras", donde sus decisiones no pueden ser explicadas fácilmente. Esta opacidad plantea problemas relacionados con la confianza.

(c) Privacidad y seguridad de los datos. La IA al depender de grandes volúmenes de datos, de muy diversas fuentes, a menudo no controladas ni validadas, plantea riesgos significativos en términos de privacidad y seguridad. De hecho, constituye un gran desafío para la protección de datos, ya que, de hecho, se dan recolecciones masivas de datos sin consentimiento, con una falta real de mecanismos efectivos de anonimato, unido a los riesgos y brechas de ciberseguridad y fugas de información.

(d) La IA puede ser utilizada para fines malintencionados, como la guerra cibernética o la manipulación de la opinión pública mediante la creación de noticias falsas, resultados de encuestas falsas, la difamación mediante imágenes manipuladas[36], etc.

(e) La irrupción de la IA hace que se esté perdiendo la confianza informática y se pierda el interés por las noticias. Los medios llevan algún tiempo utilizándola y en gran medida "detrás de escena". Aunque el público está aprendiendo sobre la IA en general y va formando sus opiniones, pocos entenderán cómo se emplean específicamente estas tecnologías en el periodismo. Es necesario, desde la ética de los informadores, recuperar la confianza informativa; se necesita más que nunca un periodismo de calidad, independiente y plural y transparente[37].

(f) La IA requiere infraestructuras muy costosas, un consumo energético muy elevado que impacta con el medio ambiente.

(g) Recientemente han aparecido numerosos casos en los que sistemas de inteligencia artificial se comportaban de forma racista o sexista. Dado que los sistemas inteligentes son creados por personas, parece inevitable que los sesgos humanos se transmitan a los propios algoritmos o a través de los datos.

(h) La interacción temprana con sistemas de IA puede influir en el desarrollo cognitivo de los niños, impactando su capacidad de resolución de problemas, pensamiento crítico y habilidades sociales[38,39,40,41,42].

Por todo ello, la ética exige que, a nivel mundial, se establezcan regulaciones que implique un control, así como medidas para paliar las desigualdades que se generan.

### *6.2. Consideraciones éticas de los usuarios para evitar una posible deshumanización*

(a) El peligro de la IA radica en deshumanizarnos con su uso hasta el punto de vernos como complicados robots vivos, supermáquinas y al mismo tiempo "antropomorfizar" la IA acostumbrándonos a usar conceptos y términos exclusivamente asociados a cualidades exclusivamente humanas.

(b) La dependencia excesiva de los sistemas de IA puede conducir a una pérdida de creatividad, capacidad de pensamiento crítico, intuición, etc. Un uso ético nos exigirá potenciar la serie de valores humanos que el uso abusivo de la IA desgasta. Si afirmamos radicalmente la

36 Kalpokas, I. and Kalpokiene, J. (2022). Deepfakes. A Realistic Assessment of Potentials, Risks, and Policy Regulation. SpringerBriefs in Political Science (BRIEFSPOLITICAL).

37 Vara Miguel, A. (2024) Calidad periodística y pluralidad: claves para la confianza informativa en la era de la Inteligencia Artificial (IA) Informe DIGITAL NEWS REPORT 2024. Universidad de Navarra. https://www.unav.edu/web/digital-news-report/entradas/-/blogs/informe-ejecutivo. Accedido, 1 de mayo de 2025.

38 Bai, L. et al. (2023). ChatGPT: The cognitive effects on learning and memory. Brain-X. 1: e30. https://doi.org/10.1002/brx2.30

39 Schemmer M. et al. (2022). Should I follow AI-based advice? Measuring appropriate reliance in human-AI decision-making. arXiv preprint arXiv. https://doi.org/10.48550/arXiv.2204.06916

40 Pedro F. et al. (2019). Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. United Nations Educational, Scientific and Cultural Organization; https://hdl.handle.net/20.500.12799/6533

41 Buçinca, Z,et al. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proc ACM Hum Comput Interact. 5(CSCW1): 88. https://doi.org/10.1145/3449287

42 Zhou, J. et al. (2023). Ethical ChatGPT: concerns, challenges, and commandments. arXiv preprint. https://doi.org/10.48550/arXiv.2305.1064. Accedido, 5 de abril de 2025.

**Tabla 3**. Comparativa rápida: España, UE y EE. UU

| Aspecto | Unión Europea | España | EE. UU. |
|---|---|---|---|
| **Ley específica** | Sí (AI Act) | Sí (aplica AI Act) | No (iniciativas y normas sectoriales) |
| **Supervisión** | Comisión Europea / AESIA | AESIA (Agencia Española de Supervisión) | NAIIO y agencias federales estatales |
| **Enfoque principal** | Basado en el riesgo | Ética, igualdad y transparencia | Seguridad nacional e innovación |
| **Protección de datos** | GDPR | GDPR | Varias leyes estatales (no federal) |
| **Marcos voluntarios** | Código de conducta IA generativa | ENIA, Carta de Derechos Digitales | NIST Framework, AI Bill of Rights |

libertad de cada ser humano, será más fácil no caer en tal reduccionismo.

(c) La creciente dependencia de la comunicación y las interacciones impulsadas por la IA podría conducir a una disminución de la *capacidad de conectar con los demás, la* empatía, las habilidades sociales. Para preservar la esencia de nuestra naturaleza social, es preciso mantener un equilibrio entre la tecnología y las relaciones humanas.

(d) A medida que los sistemas de IA se integran cada vez más en nuestra vida diaria, surge la posibilidad de pérdida de autonomía en la toma de decisiones. Desde qué leer hasta qué ruta tomar, muchos usuarios delegan decisiones en asistentes virtuales o algoritmos de recomendación sin *cuestionar sus sugerencias*.

(e) El uso de las herramientas como *ChatGPT* exige juicio crítico. A veces, especialmente cuando se trabaja con referencias muy específicas y concretas (como citas académicas, libros o artículos muy determinados), la herramienta "se inventa" datos o referencias bibliográficas. Esto se llama *alucinación* y es una limitación conocida de los modelos. Se puede dar especialmente cuando se trata de temas poco documentados o muy recientes, o cuando el usuario, especialmente si es novel en el manejo de estas herramientas no especifica una fuente concreta. La respuesta generada puede ser un ejemplo plausible pero ficticio. Es necesario un análisis profundo de la información facilitada por la herramienta, para detectar estas alucinaciones, posibles sesgos, además de comprobar exhaustivamente las referencias bibliográficas aportadas, su selección cuidadosa, tanto para no cometer errores como para respetar los derechos de autor y la propiedad intelectual, de autores no citados, pero de los que se ha obtenido la información[43].

## 7. Regulación de la inteligencia artificial en la actualidad

La regulación de la IA está evolucionando rápidamente para adaptarse a su creciente presencia en todos los sectores. Los marcos normativos varían mucho, si bien comparten principios comunes como la protección de los derechos fundamentales, la transparencia, la rendición de cuentas y la seguridad.

En la tabla 3 (Generada con ChatGPT: GPT-4o mini) se incluye una breve comparativa.

En la Unión Europea (UE), está vigente la Ley de Inteligencia Artificial de la UE (AI Act), que fue aprobada en 2024[44], la primera gran legislación integral sobre IA del mundo. Su aplicación será gradual en el binomio 2025-2026. Su enfoque está basado en el riesgo.

Hay normativas complementarias como el Reglamento General de Protección de Datos (GDPR, por sus siglas en inglés), que protege la privacidad y regula el uso de datos en IA. También se aplica la Carta de Derechos Digitales de la UE que orienta el desarrollo ético y equitativo de la IA, o el Código de Conducta Voluntario para Modelos Generativos (como GPT o DALL·E): que establece buenas prácticas mientras se implementa el AI Act.

43 López Moratalla, N. "Inteligencia Artificial ¿Conciencia artificial? Una perspectiva desde las ciencias de la vida" Madrid, Digital Reasons. 2017.

44 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)

En España, rigen las leyes de la UE y además se aplica la llamada Estrategia Nacional de Inteligencia Artificial (ENIA) establecida en 2020, que define la hoja de ruta para el desarrollo y uso ético de la IA. Establece principios como la igualdad, la transparencia y la sostenibilidad.

En 2024 se creó la Agencia Española de Supervisión de la IA (AESIA), con sede en La Coruña, ya operativa desde 2024, que es el primer organismo europeo de supervisión dedicado exclusivamente a la IA. Se encargará de supervisar la aplicación del AI Act en España y garantizará el cumplimiento ético de la IA en el sector público y privado. En julio de 2021 el Ministerio de Transformación Digital de España publica la Carta de Derechos Digitales[45].

En Estados Unidos (EE. UU.), aún no hay una ley federal específica sobre IA, pero ha establecido un conjunto de iniciativas y regulaciones parciales, como la US Executive Order on Safe, Secure, and Trustworthy AI (octubre 2023)[46], una orden ejecutiva firmada por la Casa Blanca, que exige evaluaciones de seguridad, transparencia en modelos generativos y protección contra sesgos. Pone énfasis en IA segura para sectores críticos: salud, defensa, empleo y educación.

También existe el NIST AI Risk Management Framework (2023), un marco voluntario creado por el Instituto Nacional de Estándares y Tecnología que ayuda a las organizaciones a diseñar sistemas de IA seguros, responsables y explicables.

Otras entidades internacionales, como la OCDE[47] han dictado unas Recomendaciones sobre una IA centrada en el ser humano (2019), pretendiendo promover principios de transparencia, seguridad y respeto a los derechos humanos.

También la UNESCO[48], confeccionó un Marco ético mundial (2021) que destaca la importancia de la equidad, la inclusión y la diversidad en la IA.

Los grandes grupos económicos y políticos de países como el G7 o el G20 abogan por el desarrollo de IA confiable, especialmente en relación con los modelos fundacionales y la gobernanza internacional.

En cualquier caso, y más allá de lo que la ley regule, está claro que, para mitigar el impacto negativo de la IA, es importante tomar medidas que garanticen su control y su uso responsable y ético, fomentando la transparencia y la responsabilidad en su diseño desarrollando políticas y regulaciones que protejan especialmente los derechos y la privacidad de los usuarios.

En definitiva, en el presente artículo abogamos por una regulación ética de los desarrollos de la IA y de su uso.

## 9. Conclusiones

La inteligencia artificial está transformando radicalmente nuestra relación con el conocimiento, el trabajo, la toma de decisiones y la relación con los demás, lo que obliga a repensar el papel de la mente humana frente a estas nuevas formas de inteligencia no biológica. A pesar de sus capacidades superiores en velocidad, cálculo y procesamiento de datos, y a pesar de que la IA podría alcanzar algunas de las cualidades intrínsecas al pensamiento humano como intuición, pensamiento crítico, empatía, juicio ético y comprensión contextual, no por ello podamos decir que la IA es pensamiento humano. Esta diferencia esencial plantea el reto de diseñar sistemas que complementen y potencien las capacidades humanas, sin reemplazarlas ni reducirlas y por parte de los usuarios no ceder al desarrollo de las capacidades humanas por dejarse sustituir por la IA.

Entre las ventajas del uso de la IA se encuentran la mejora en el acceso al conocimiento, la eficiencia en la gestión de recursos y la capacidad de personalizar servicios en ámbitos como la educación o la salud. Sin embargo, también emergen efectos negativos preocupantes, como la automatización deshumanizante, la manipulación informativa, o la delegación de decisiones complejas en sistemas opacos.

Frente a estos desafíos, es imperativo desarrollar marcos éticos y normativos sólidos que garanticen la transpa-

45 Gobierno de España, Derechos Digitales. Plan de Recuperación, Transformación y Resiliencia (2021) https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf. Accedido, 2 de agosto de 2025

46 Federal Register. Executive Order 14110 of October 30, 2023. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Accedido, 6 de abril de 2025.

47 OCDE. AI Principles. https://www.oecd.org/en/topics/sub-issues/ai-principles.html. Accedido, 6 de abril de 2025.

48 UNESCO. https://www.unesco.org/es/artificial-intelligence/recommendation-ethics. Accedido, 6 de abril de 2025.

rencia, la equidad, la rendición de cuentas y la preservación de los derechos humanos. El futuro de la IA no debe medirse solo por su potencia técnica, sino por su capacidad de integrarse éticamente en sociedades libres y justas.

La educación es clave para adaptarse al nuevo paradigma tecnológico.

## Contribución de los autores

Natalia López-Moratalla: Idea original y Dirección. Corrección y remodelación texto original. Carmen de la Fuente: crítica y análisis del texto original. María Font: Redacción del texto original, búsqueda y análisis de bibliografía seleccionada. Correcciones y formato final.

## Conflicto de intereses

Las autoras declaran que no existen conflictos de interés

## Referencias

Ahmad, Z. (2023) Artificial Intelligence or Augmented Intelligence? International Journal of Science and Research (IJSR). 12: 1782-1788. https://dx.doi.org/10.21275/SR231212220052.

Amat-Rodrigo, J. (2017) Máquinas de Vector Soporte (Support Vector Machines, SVMs) disponible con licencia CC BY-NC-SA 4.0 en https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines. Accedido, 4 de abril de 2025.

Aparicio-Gómez, WO. and Aparicio-Gómez, OY. (2024). Principios éticos para el uso de la Inteligencia Artificial Ethical principles for the use of Artificial Intelligence. Revista Internacional de Desarrollo Humano y Sostenibilidad 71.1.

Bai, L. et al. (2023). ChatGPT: The cognitive effects on learning and memory. Brain-X. 1: e30. https://doi.org/10.1002/brx2.30.

Baidoo-Anu, D. and Owusu, AL. (2023). Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. SSRN. https://doi.org/10.2139/ssrn.4337484.

Buçinca, Z. et al. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proc ACM Hum Comput Interact. 5(CSCW1): 88. https://doi.org/10.1145/3449287.

Celeghin, A. et al. (2023). Convolutional neural networks for vision neuroscience: significance, developments, and outstanding issues. Frontiers in Computational Neuroscience .17. DOI:10.3389/fncom.2023.1153572.

Copeland, J. (2003). The Turing Test, en Moor Moor, James James, ed., The Turing Test: The Elusive Standard of Artificial Intelligence (Springer), ISBN 1-4020-1205-5.

Corrêa, NK. et al. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns. 4: 100857. https://doi.org/10.1016/j.patter.2023.100857.

Dergaa, I. et al. (2024). Using artificial intelligence for exercise prescription in personalised health promotion: a critical evaluation of OpenAI's GPT-4 model. Biol. Sport 41: 221–241. DOI: 10.5114/biolsport.2024.133661.

Federal Register. Executive Order 14110 of October 30, 2023. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Accedido, 6 de abril de 2025.

Garrido, EC. and Lumbreras, S. (2022). On the independence between phenomenal consciousness and computational intelligence. https://doi.org/10.48550/arXiv.2208.02187. Accedido, 6 de abril de 2025.

Genova, G. (2025) Aceptado el primer artículo científico generado por IA. https://theconversation.com/aceptado-el-primer-articulo-cientifico-generado-por-ia-253451 Accedido, 29 de abril de 2025.

Goodfellow, IJ. et al. (2014). Generative Adversarial Networks. https://doi.org/10.48550/arXiv.1406.2661. Accedido, 4 de abril de 2025.

Gobierno de España, Derechos Digitales. Plan de Recuperación, Transformación y Resiliencia (2021) https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf. Accedido, 2 de agosto de 2025

Griffiths, TL. (2020) Understanding Human Intelligence through Human Limitations. Trends in Cognitive Sciences. https://doi.org/10.48550/arXiv.2009.14050.

Honton, G. et al. (2015) Deep learning. Nature, 521: 436-444. https://doi.org/10.1038/nature14539.

Hutton, JS. et al. (2020). Associations Between Screen-Based Media Use and Brain White Matter Integrity in Preschool-Aged Children. JAMA Pediatrics. 174(1):e193869. doi:10.1001/jamapediatrics.2019.3869.

Inglada, L. et al. (2024) Ethics and artificial intelligence. Revista Clínica Española (English Edition), 224: 178-186.

Kalpokas, I. and Kalpokiene, J. (2022). Deepfakes. A Realistic Assessment of Potentials, Risks, and Policy Regulation. SpringerBriefs in Political Science (BRIEFSPOLITICAL).

Kasneci, E. et al (2023) ChatGPT for good? On opportunities and challenges of large language models for education. Learn Indiv Differ. 103:102274. https://doi.org/10.1016/j.lindif.2023.102274.

Khan, AA. et al. Ethics of AI: A Systematic Literature Review of Principles and Challenges. arXiv:2109.07906. https://doi.org/10.48550/arXiv.2109.07906. Accedido, 6 de abril de 2025.

Lindsay, RK. et al. (1980). Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project. McGraw-Hill Book Company, 1980.

López-Moratalla, N. (2017) Un puente mente cerebro. Ideas 8, 33-45.

López-Moratalla, N. (2017). "Inteligencia Artificial ¿Conciencia artificial? Una perspectiva desde las ciencias de la vida" Madrid, Digital Reasons. 2017.

López-Moratalla, N. (2020) La neurología actual en el origen de lo humano. NATURALEZA Y LIBERTAD. Revista de estudios interdisciplinares. 13, 87-101.

López-Moratalla, N. (2021) en: Humanos. Vínculos familiares en el corazón del cerebro. Ed RIALP, pp. 17-23.

Marciano, L. et al. (2021). The Developing Brain in the Digital Era: A Scoping Review of Structural and Functional Correlates of Screen Time in Adolescence. Front Psychol. 12:671817. doi: 10.3389/fpsyg.2021.671817.

McCarthy, J. et al. (1955). A Proposal for The Dartmouth Summer Research Project On Artificial Intelligence. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html. Accedido, 2 de abril de 2025.

Mcculloch, WS. and Pitts, W. (1990). A logical calculus of the ideas immanent in nervous activity Bulletin of Mathematical Biology. 52: 99-115.

Monroe, D. (2014). Neuromorphic computing gets ready for the (really) big time. Communications of the ACM. 57: 13 – 15. https://doi.org/10.1145/2601069.

Nivins, S. et al. (2024). Long-term impact of digital media on brain development in children. Sci Rep. 14:13030. doi: 10.1038/s41598-024-63566-y.

OCDE. AI Principles. https://www.oecd.org/en/topics/sub-issues/ai-principles.html. Accedido, 6 de abril de 2025.

Oficina de Prensa de la Santa Sede. (2025). Nota sobre la relación entre la inteligencia artificial y la inteligencia humana. Vatican News. Accedido, 6 de abril de 2025.

Pedro, F. et al. (2019). Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. United Nations Educational, Scientific and Cultural Organization; https://hdl.handle.net/20.500.12799/6533.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance).

Rosenblatt, F. (1958). Perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review. 65: 386-408. Accedido, 2 de abril de 2025.

Rumelhart, D. et al. (1986) Learning representations by back-propagating errors. Nature. 323: 533–536. https://doi.org/10.1038/323533a0.

Sakana IA (2024) The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. https://sakana.ai/ai-scientist/. Accedido, 29 de abril de 2025.

Sánchez, C. (2025). Por qué no podemos afirmar que la inteligencia artificial 'habla'. The Conversation.

https://theconversation.com/por-que-no-podemos-afirmar-que-la-inteligencia-artificial-habla-250096. Accedido, 30 de abril de 2025.

Sánchez Martín, FM. et al. (2007). History of robotics: from archytas of tarentum until da Vinci robot. (Part I). Actas Urológicas Españolas. 31: 69-76. https://doi.org/10.1016/S0210-4806(07)73602-1.

Schemmer M. et al. (2022). Should I follow AI-based advice? Measuring appropriate reliance in human-AI decision-making. arXiv preprint arXiv. https://doi.org/10.48550/arXiv.2204.06916.

Shortliffe, EH. (1976). Computer Based Medical Consultations: MYCIN, American Elsevier, 1976.

Sutton, RS. and Barto, AG: (2018) Reinforcement Learning: An Introduction MIT Press, Cambridge, MA Second Edition.

Toolify.ai (2024) Aprendizaje Auto-Supervisado: Una Nueva Frontera en IA. https://www.toolify.ai/es/ai-news-es/aprendizaje-autosupervisado-una-nueva-frontera-en-ia-1766168. Accedido, 4 de abril de 2025.

Turing, A. (1948). Machine Intelligence, en Copeland, B. Jack, ed., The Essential Turing: The ideas that gave birth to the computer age, Oxford: Oxford University Press, ISBN 0-19-825080-0.

UNESCO. https://www.unesco.org/es/artificial-intelligence/recommendation-ethics. Accedido, 6 de abril de 2025.

Vara Miguel, A. (2024) Calidad periodística y pluralidad: claves para la confianza informativa en la era de la Inteligencia Artificial (IA) Informe DIGITAL NEWS REPORT 2024. Universidad de Navarra. https://www.unav.edu/web/digital-news-report/entradas/-/blogs/informe-ejecutivo. Accedido, 1 de mayo de 2025.

Vaswani, A., et al. (2017). Attention is All You Need. https://arxiv.org/abs/1706.03762. Accedido, 4 de abril de 2025.

Yax, N. et al. (2023) Studying and improving reasoning in humans and machines. arXiv:2309.12485. Accedido, 5 de abril de 2025.

Zhou, J. et al. (2023). Ethical ChatGPT: concerns, challenges, and commandments. arXiv preprint. https://doi.org/10.48550/arXiv.2305.1064. Accedido, 5 de abril de 2025.