E.M. nº 79 Enero-Abril 2025

INTELIGENCIA ARTIFICIAL: ¿UNA QUIMERA?

Isabel De Val Pardo
Catedrático de Organización de Empresas

RESUMEN

¿La IA puede emular en "sentido humano/modo humano" la conciencia cuando se trata de una experiencia subjetiva en primera persona? De momento la IA no comprende el lenguaje ni tiene consciencia: la comprensión, intencionalidad y conciencia son intrínsecamente humanas y no pueden reducirse a procesos computacionales/mecánicos. El reto es simular cerebro, cuerpo y mente en distintos soportes y agentes corpóreos que perciban, razonen, actúen, que sean conscientes ¿Se puede replicar el origen natural/humano de la inteligencia, mente, consciencia e intelecto de manera artificial? La toma de decisiones humanas deriva al utilizar sistemas no biológicos cotidianamente ¿estamos seguros de las que toman los algoritmos y afectan nuestras vidas?

1 INTRODUCCIÓN

El conocimiento de la fisiología humana transita de la visión mecánica de los seres vivos (tradición dualista) que propone la separación de mente y cuerpo (según Descartes el cuerpo prescinde de pensamiento al estar compuesto por piezas mecánicas que realizan una funciones), a la concepción de los humanos como agentes cognitivos que logran lo que hacen vía cuerpo, cerebro, mente, sistema nervioso periférico y entérico insertos en un mundo en el que proceso vital no es lo mismo que inteligencia (no exclusiva del cerebro, equívoco que origina la similitud con una computadora).

La metáfora cerebro-computador sugiere que los estados y procesos mentales son procesos computacionales del hardware neuronal del cerebro, incluido el razonamiento: paralelismo que utiliza una posición intencional para predecir y explicar el comportamiento de una entidad al tratarla como si tuviera intenciones, creencias y deseos. El atribuir intencionalidad de explicación y predicción es clave (La Fontaine, 2024) en cuanto a la naturaleza de la conciencia y la mente (se entienden como procesos algorítmicos sujetos al mismo análisis que los programas de una computadora), puede ser útil para comprender sistemas complejos y remite al laberinto de la inteligencia.

Algo es inteligente si tiene habilidad de percibir información, la retiene como conocimiento y aplica a comportamientos que se adaptan a entornos cambiantes (Kaspar *et al.*, 2021: 345-347). Las materias artificiales calificadas de inteligentes se inspiran en la naturaleza, aprenden de organismos vivos y se adaptan al combinar elementos funcionales esenciales que en el logro de un cometido lo sintetiza -según complejidad y funcionalidad- la interacción de un sensor, un agente, una red y una memoria si bien carecen de la materia sintética propia de los humanos, la cognición.

La inteligencia "*desde la perspectiva general de los organismos vivos se refiere a la capacidad de resolver con éxito los problemas planteados por la lucha de la vida*" (Damasio, 2021:45) y el referente de significado es la humana que según Bridle (2024) se manifiesta en cualidades como la capacidad de pensar de manera lógica, comprender, autoconciencia, aprendizaje, comprensión

emocional, creatividad, razonamiento, resolución de problemas y planificación; todas ellas precisan de consciencia, mente y sentimientos.

A pesar de todo lo que se ignora del cerebro muchas ciencias profundizan en su conocimiento y tratan de avanzar vía ingeniería inversa (para comprender cómo funciona un ordenador se recurre a redes neuronales y para conocer cómo funciona el cerebro se recurre al ordenador) incluida la naturaleza de la mente; también se persigue computación neuromórfica descentralizada que simule la inteligencia humana (IH) sin comprensión ni posesión de estados mentales/intencionales; y máquinas que realicen tareas donde se aplique la inteligencia, el razonamiento y el conocimiento humano. La inteligencia artificial **(**IA) definida como "el proceso de hacer que una máquina se comporte de forma que sería llamada inteligente si un ser humano la hiciera" (Castilla, 2023) se debe a la interrelación de ciencias cognitivas y de computación, y aglutina tecnologías que simulan procesos cognitivos humanos de razonamiento, aprendizaje y autocorrección soportados en instrumentos materiales vía computación, máquinas y artefactos con el uso de algoritmos, tecnologías y aplicaciones cognitivas, robóticas o interfaces.

Y surgen ciertas cuestiones ¿inteligencia implica consciencia? ¿Es prerrequisito? ¿La consciencia es condición necesaria para hablar de inteligencia? La consciencia (incluye pensamiento, sentimiento y percepción) es un fenómeno exclusivamente humano, no surge por sí solo en el cerebro, el cual no es suficiente de manera aislada (*per se* no obtiene vida consciente cual *brain in vat*), no puede emerger en máquinas que no estén constituidas por células que dependan de oxígeno y nutrientes, que no estén vivas en sentido biológico (Narby, 2023) y adolezcan de la capacidad de replicarse. La diatriba reside en "aspecto vital *versus* material inerte"; de nuevo la similitud artificial-biológico busca argumentos, formulas, ecuaciones y tecnologías que alienten vida y las opiniones difieren en cuanto a si una máquina es consciente, "*algo más que permita maximizar una recompensa*" (Seth, 2021: 297).

Los sistemas de IA son inteligentes en sentido no significativo, reconocen patrones y hacen lo que hacen sin ser conscientes de nada; de manera similar se usa el término conciencia si la IA responde a estímulos, aprende algo o alcanza un objetivo. Si en organismos vivos existe inteligencia sin conciencia en lo artificial se puede identificar por medio del adecuado proceso de información (Tsuchiya, 2017). En particular la Teoría de la Información Integrada (TII) afirma "*que la conciencia no es más que información integrada producida por un sistema totalmente determinado por propiedades de sus mecanismos internos, por su estructura de causas-efectos*" (Seth, 2021: 300), una perspectiva entre las numerosas que tratan la conciencia en la IA (Butlin y Long, 2023); en algunas es sinónimo de consciencia y difieren en las que se sustentan en la Teoría de la Mente (Farhadi, 2025), es el caso de la Trilogy Theory of Consciousness (TTC) al poner de relieve las áreas críticas entre IH e IA que se concretan en la intención, autoconciencia y agencia.

2 DE TELARES JACQUARD A MODELOS DE LENGUAJE

Aludir los intentos de simular el comportamiento humano por medio de criaturas artificiales animadas, artefactos, instrumentos y soportes técnicos se concretan en el talento, conocimiento, genialidad y singularidad de quienes protagonizan el fugaz tránsito que va del "telar al modelo lingüístico" entre ciencias, máquinas, autómatas, el "juego de imitación" o la "habitación china", incluso ficción literaria y cinematográfica. De manera sucinta el texto (Mary Shelley y Fritz Lang en la mente) se vale de la industria textil, de codificar las instrucciones de la máquina analítica de Babbage y alude a conocimientos que en la actualidad alcanzan a un amplio espectro de procesos y actividades de distintos sectores económicos e inciden en la sociedad general.

- ✓ La era de la información -la distinción entre datos y procesamiento- la vislumbra Ada Lovelace (Essinger, 2014) al comprender las consecuencias de la máquina analítica de Babbage y aplicarlas al telar de Jacquard. Se considera el antecedente del dígito binario (es el primer ejemplo en la historia de la tecnología humana del proceso de digitalización con

tarjetas perforadas) que se podría aplicar por extensión a todo proceso que implicara tratar datos. Según Ada Lovelace las máquinas no pueden definirse "inteligentes", no tienen intención y sólo hacen lo dispuesto por el constructor que al pensar de antemano determina la acción a realizar para el trabajo encomendado ya que la finalidad requiere una serie de actos tendiendo a conseguir el objeto determinado (en la actualidad los humanos son dueños de los algoritmos que programan un desempeño productivo).

- ✓ El nexo de IA y robótica lo anticipa Torres Quevedo ya que un "autómata con discernimiento" debe ser capaz de conocer las circunstancias del entorno que le permitan desarrollar una conducta de adaptación al mismo. Considera que los autómatas debían disponer de sentidos que les suministren noticias sensoriales del mundo, de sus miembros, de energía que posibiliten movimientos y acciones exteriores; y apunta la necesidad de imitar a los seres vivos al tratar la idea de máquinas con capacidad de decisión, con vida de relación que se orienten a la simulación de la acción humana, vía Automática, ciencia de la que fue pionero entre tantos campos en los que despuntó, incluida la robótica y el diseño de un escenario mecánico cual "habitación china" de Searle (Garrido 2003; Blanco Pérez, 2019).

- ✓ El juego de imitación o test de Turing se orienta a la simulación automática del pensamiento humano, de actividad mental por procedimientos no biológicos y aunque no concreta los detalles del mismo avanza la presencia de inteligencia, predice la posibilidad de que la IA adquiera capacidad para programarse, de cognición a nivel humano y su alineación con los humanos en cuanto a lenguaje y razonamiento con sentido común. En 1950 el test del juego inspira las capacidades de la IA y cuestiona si en una conversación "una máquina puede pasar con éxito por una conciencia humana inteligente" cuando recibe información, hace algo con ella sin ser consciente de nada y actúa sin pensar.

Detectar cognición de nivel humano por medio del test determina lo que se conoce como "máquina universal" (base de estandarización para la mayor parte de ordenadores al averiguar si el interlocutor es humano o mecánico en función de cómo responde al sentido de la respuesta) que se limita al recuento del reordenamiento de símbolos o piezas de cálculo sin suministrar información sobre el mundo real, ni lo puede hacer por la imposibilidad de establecer contacto con el exterior. Se trata de una máquina sintáctica que en una conversación responde de manera indistinguible de un humano (Carrera, 2024), hace lo que se le ordena al tener un propósito explícito y no se adapta a situaciones imprevistas al estar las operaciones limitadas a la información introducida.

- ✓ La metáfora de la habitación china de Searle ocasiona mera apariencia de inteligencia, mente y pensamiento. En su interior un sujeto que no sabe chino ni comprensión semántica, dispone de unas reglas con las que puede manipular ideogramas chinos, recibe secuencias de símbolos a transformar y construir otras nuevas a trasladar al exterior según las normas dadas, de aquí la aparente conversación inteligente de quién supiera chino. De manera similar los algoritmos procesan el lenguaje humano sin comprender el texto, detectan patrones y devuelven lo más probable en base a asociaciones estadísticas en textos analizados: cuando *Google Translate* recibe un input, lo procesa siguiendo unas reglas, completa su tarea mecánica e inconsciente y emite un output, lo que ejemplifica la manipulación de la sintaxis (los símbolos) sin comprensión del lenguaje (significado), ni tener conciencia.

Comprender el lenguaje no sólo consiste en realizar una tarea de tipo simbólico, requiere tener una interpretación o un significado unido a dichos símbolos de la que una máquina carece (Santos Sousa, 2005), pues los programas no tienen mente ni relación causal y la simulación plena de fenómenos mentales por medios técnicos carece de experiencia subjetiva (Jurado González, 2023; Morandín-Ahuerma, 2023).

- ✓ Los modelos de lenguaje (LMs), jocosamente calificados de "loros estocásticos" (Bender *et al.*, 2021) son útiles y cuestionables (Arkoudas, 2023; La Fontaine, 2024) desde la mera atribución de "inteligentes" sin comprender lo que es inteligencia. Calculan las probabilidades de que una palabra aparezca después de otra en el lenguaje humano, así generan texto, dicen sin entender ni saber lo que dicen (son "habitaciones chinas") y aunque se denominen modelos de lenguaje pretenden modelizar; no es lenguaje humano entendido como capacidad cognitiva (sistema de conocimiento o facultad que permite aprender y usar una lengua cualquiera) sino la probabilidad de aparición de grupos de caracteres basándose en las frecuencias observadas en los textos de entrenamiento.

En particular ChatGPT es un modelo matemático producto del uso del lenguaje humano, no de la capacidad de producir el lenguaje humano. Es un sistema de lengua externa y extensional (lengua-e) suma de todas las oraciones producidas por los hablantes de una lengua dada, es decir constructo sociocultural que pasa de generación en generación; y al crear secuencias de caracteres a partir de otras es "*un loro estocástico en una habitación china"* (Mendivil, 2023: 10-21). Establece correlaciones del cúmulo disponible de términos y aunque ni sepa lo que es un humano habla como tal, las respuestas son coherentes, relevantes y entendibles sin saber ni comprender lo que dice: presenta ausencia total de intelecto.

La IA generativa tiene dominio en capacidad de información, pero no en explicación de las interconexiones que se dan vía estadística y probabilidad, pero la limitada capacidad de comprensión del lenguaje, aunque haya superado el test de Turing (¿mera verdad cuantificable de lo artificial como el coeficiente intelectual humano para "*cómo ser*"?) en una proporción reducida no demuestra inteligencia humana y carece de lo fundamental para la consciencia y conciencia subjetiva.

- ✓ Asimov anexiona ciencia y ficción a partir de las leyes de robótica (un robot no dañará a un humano, ni por inacción permitirá que sufra daño; cumplirá órdenes dadas por los humanos salvo si entran en conflicto con lo anterior; protegerá su propia existencia mientras que no entre en conflicto con todo lo previo: Rocha, 2024) al apuntar, en un escenario ficticio, la visión ideal de la interacción humano-máquina que impregnan el pensamiento sobre ética de la IA. Los robots bajo los imperativos de las tres leyes se comportan moralmente correctos, pero podrían no hacerlo: se asimilan a los humanos, unos y otros pueden llegar a incumplirlas (caso de Nathan, Caleb y Ava en *Ex Machina*); a pesar de ser genéricas y universales se han actualizado y ampliado (Pasquale, 2024) a fin de atender el devenir difuso y complejo con interferencias políticas, económicas y sociales.

- ✓ En 1968 Arthur Clarke (*2001: A Space Od*yssey) anticipa una utopía, pura tecnología humanoide -HAL 9000- que entraña inteligencia artificial no corpórea. En espera del progreso científico y tecnológico que ofrezca modelos no corpóreos cabe preguntarse qué se le ocurriría a Bowman al traspasar puertas lógicas, si la computación cuántica le facilitara proseguir la andadura, dada la conciencia que vislumbra el comportamiento de HAL ¿Atenderían las leyes de Asimov? Cuestión que remite a la no subjetividad de la robótica e IA, a la carencia de cerebro, cuerpo y mente, a la conciencia/consciencia como prerrequisito de inteligencia general.

- ✓ En la ficción cinematográfica Alex Garland (*Ex Machina*, 2014) transforma el test de Turing en uno de consciencia: un agente con cuerpo, Ava robot inteligente, trata de confundir a Caleb para hacerle creer que es consciente, que su comportamiento es indistinguible de un humano con propósitos, deseos y temores de su fin ¿será destruida? (los test de Turing y Garland no miden si las máquinas piensan o no, tratan si los humanos lo hacen al presuponer que son inteligentes, conscientes y agudos).

3 ¿CONCIENCIA ARTIFICIAL?

La definición en distintas ciencias -filosofía, religión, neurociencia, psicología- difiere y confunde, es "terreno minado" -según Damasio- al que se incorporan la física, ingeniería, computación y robótica en el intento de resolver si máquinas o IA pueden pensar, ser inteligentes y conscientes en la toma de decisiones al interactuar con el entorno. Si la naturaleza de la conciencia es biológica una máquina no podrá simular plenamente procesos mentales y si no se sabe cómo la consciencia se genera en la mente humana en la que se aglutinan procesos subconscientes, inconscientes, sentimientos, intuición, imaginación, inspiración, creatividad, inteligencia y aprendizaje ¿cómo la IA puede tener autoconciencia sin experiencia subjetiva?

La mayor parte de las funciones del cerebro pueden explicarse mediante leyes físicas y químicas dentro de la neurofisiología y pueden reproducirse en un ordenador o prótesis biónicas, pero no pueden procesar sentimientos ni añoranzas (Campillo, 2021). La consciencia, en ausencia de prueba científica que la confirme o niegue, es una cuestión filosófica con bases biológicas y culturales que realimentan distintas disciplinas. La utilización conveniente de "inteligente e inteligencia" origina que al nivel de complejidad y organización en el procesamiento de información se entienda como "conciencia en sentido humano"; a mayor abundamiento la contribución de la evolución exponencial de la IA generaliza su extensión y difumina la naturaleza.

Se alude a la conciencia (Kurzweil, 2025) en el sentido de capacidad funcional que permite a los humanos percibir lo que les rodea y actuar de manera consciente simultáneamente por los pensamientos propios y del mundo exterior (en tal caso es consciencia); o de capacidad subjetiva, *qualia*, término en el campo de la filosofía de la mente que define las cualidades subjetivas de las experiencias mentales (un color, un sabor o dolor, incluso la tecnología trata de ir más allá del cerebro biológico y ampliar sus cualidades: ser inefables, intrínsecas, privadas y aprehensibles: pura fenomenología) que conforman conciencia subjetiva.

La gran incógnita de la conciencia gira en torno al por qué, cómo y dónde se construye, lo que entraña el conocido *hard problem* (Chalmers, 1995, 2018), barrera insuperable o *gap* explicativo (Herzog y Herzog, 2024; Tiuninas, 2025) para la IA, sin solución universal, depender de distintas perspectivas, teorías y tipos de modelos (ej.: Smith y Schillaci, 2021; Chalmers, 2023; Butlin y Long, 2023; Farhadi, 2025). El por qué alude a los mecanismos complejos descifrables que facilitan al cerebro la construcción de imágenes e instrumentos de manipulación vía mapas (memoria, lenguaje, razonamiento y toma de decisiones); el cómo se refiere a la construcción de la experiencia mental y los sentimientos que acompañan a las imágenes (Damasio, 2018); y el dónde los procesos tienen lugar, en los que intervienen elementos visuales, auditivos u olfativos y la integración de experiencias. ¿Cómo se transforma la información en experiencias subjetivas y las sensaciones en percepciones? ¿A través de qué mecanismos? Se trata de un modo de *ser* propio de algún modo (Nagel, 1974).

La consciencia en sistemas físicos se inspira en el cerebro, esto requiere modelos relevantes que permitan cómo fusionar la información que seleccionan y el propio control de veracidad (Dehaene *et al.*, 2021) en el marco de la neurociencia cognitiva según la cual: el prerrequisito de la consciencia es la atención, el proceso cognitivo es factible sin consciencia y ésta es imprescindible para las operaciones mentales (Dehaene y Naccache, 2001) y la única teoría que distingue entre conciencia y consciencia es la TTC (Farhadi, 2025: 17-18) por ser la primera prerrequisito, de aquí que hasta que el *hard problem of consciousness or awareness* no se resuelva, la noción en la IA será mera especulación.

La IA atiende/no atiende la distinción entre conciencia (como experiencia) y consciencia (como estado de la mente) según teorías; en términos de ficción cinematográfica "consciencia artificial como secuela", haberla, ahíla y "*un sistema de IA podría considerarse consciente si se comporta de forma indistinguible a la de un ser consciente sin necesidad de atribuir estados mentales al sistema*" (Morandin-Ahuerma, 2023: 203; Tuleubekov *et al.*, 2023), pero la experiencia subjetiva es clave lo que

dificulta la extensión a la IA y cuestiona la relación mente-materia. Los modelos de IA que simulan consciencia se califican "agentes" aunque racionalidad y eficiencia no son condición necesaria de *self* (puede sean del *ego* humano) y artificialmente se pretende provocar comportamientos que simulen atribución de conciencia en computadoras, robots y sistemas de información por medio de algoritmos y sistemas de redes neuronales que identifiquen sensaciones.

El avance exponencial de la IA da lugar a robots de aspecto humano que aparentan tareas mentales fáciles para los humanos**,** pueden tener pura cognición, pero no afecto a pesar de semblantes con sentimiento. Es el caso de Ava (*Ex Máchina*), un robot humanoide que trata de identificar el juego: parece viva, programada sin emociones motivadas ni experiencias subjetivas que le faciliten una perspectiva individual de su propio organismo y de sentimientos individuales. Los creadores de la IA esperan alineación de las tareas que programan para el logro de objetivos con el buen hacer humano (de utopía/idealismo al riesgo cero no existe) cuando son ejemplo en desalinear propósitos éticos y morales en la humanidad: a pesar del amplio espectro de legislación existente los derechos y libertades no se respetan ¡*Nobody´s perfect!*

4. ADENDA

- ✓ Ada Lovelace, con una férrea educación y formación matemática en un entorno social privilegiado que la estimula intelectualmente, mantiene una relación con Babbage y fascinación por las máquinas. Gracias a sus conocimientos, intuición y visión percibe el potencial tecnológico de la máquina analítica superando las ideas del propio autor: estima que era más que una calculadora, con limitaciones ya que sólo podía hacer aquello que se le había encargado, no crea, sólo facilita lo conocido sin prever relaciones ni verdades analíticas. Esta incapacidad denominada "la objeción de Lovelace" por Turing contribuye a las bases para saber si las máquinas piensan y la creación del ordenador.

- ✓ Turing se plantea la posibilidad que las máquinas pensaran por equivocación propia o porque malentendieran su función; duda que le remitía a cuestionar la existencia de conciencia en lo artificial. Predijo que al avanzar la tecnología se podrían "*diseñar sistemas de circuitos que pudieran adaptarse a nuevos datos e información*" (Bridle, 2024: 54) y ordenadores con limites similares al pensamiento humano ya que clasificaba los organismos de cómputo en binarios: humano o no humano y su test se centra en preguntas concretas con respuestas con el que fingir inteligencia equivale a tener inteligencia, lo que cuestiona la experiencia de la "habitación china".

- ✓ La metáfora de Searle evidencia que el ordenador no es inteligente, no piensa ni tiene comprensión de cómo funciona ya que el proceso de información de una computadora no constituye una mente dada la carencia de "*dinámica con un mundo circundante*" (Noë, 2010); de aquí que distinga entre IA fuerte (una mente que piense como los humanos) y débil (la artificial que ayude a las actividades mentales humanas).

- ✓ Dado el auge de los LMs insistir que, aunque aparentan/simulan inteligencia y mente, no piensan ni entienden los elementos semánticos de una lengua: los intentos de imitación y desarrollo por parte de la computación, neurociencia y otras aproximaciones muestran indicadores de consciencia (Butlin y Long, 2023) no condiciones de consciencia (Chalmers, 2023).

El avance de la creación de una inteligencia similar a la humana se puede atribuir a Turing al apuntar que una máquina se dotara de las capacidades cognitivas de un niño (Carrera, 2024) lo que permite a Pearl y Mackenzie (2020) defender un modelo causal para que interactúe repetidamente con el mundo físico y social, aprenda y adquiera conocimiento causal al introducir información parecida a

la que se recibe en la infancia, y el programador ejerza el rol de quienes contribuyen a su evolución: la interacción podría alumbrar conciencia artificial.

Teorías o modelos causales son una esperanza para el avance cognitivo: en biología el comportamiento de un organismo vivo (De Waal, 2008) responde al "dominio de causación última" que se refiere al porqué del mismo y al "dominio de causación próxima" que alude a situaciones inmediatas y estímulos que lo han provocado; la primera ocasiona cuestionarse el "por qué" y el "para qué" y la segunda "con quién" y de "qué modo", interrogantes con los que sortear predicciones basadas en observaciones pasivas y facilitar científicamente qué es la inteligencia y la conciencia.

Mientras, el método socrático contribuye a la reflexión y búsqueda de conocimiento. ¿La consciencia es una prerrogativa humana? ¿El hardware de una computadora la puede replicar al generarse por material biológico? Las cuestiones se aglutinan y circunscriben a si lo "artificial inteligente" alcanzará las características humanas (inteligencia general y conciencia) o las superará. La naturaleza de equiparación responde a procesos inversos: lo artificial de fuera-dentro (piezas que se ensamblan y una unidad central: un ordenador/computadora/robot que no procesa lo inmaterial) y lo biológico de dentro-fuera (células eucariotas y homeóstasis: un humano con cuerpo, cerebro, mente, sistema nervioso, ciclo percepción/acción, emociones y libre albedrio).

En el caso de la mente si "*surge de la interacción de cuerpo y cerebro"* (Damasio, 2018: 274) no se puede simular, no puede descargarse en un ordenador y el sentido propio se reduce/pervierte al trasladar el sustrato de un tipo particular de química organizada de organismos vivos a meros algoritmos. Es lo que hay, pero usos "por defecto" de términos como inteligencia, conciencia, consciencia o agente confunden y penalizan el conocimiento: lo artificial llegará a lo que llegue, será distinto a lo biológico, ajeno a valores y necesidades humanas (López de Mantarás, 2015, 2018), con un código fuente por descifrar, necesidad de comprender la relación con los humanos y de entender lo que significa no ser biológico caso de mostrar características que se asocien a la vida, sin olvidar que las leyes de Asimov culminan en contradicciones.

5. REFERENCIAS BIBLIOGRÁFICAS

ARKOUDAS, K. (2023): "ChatGPT is no stochastic parrot. But it also Claims that 1 is greater than1", *Philosophy & Tehnology*, 36:54.

BENDER, E.; GEBRU, T.; McMILLAN-MAJOR, A.; SHMITCHELL, S. (2021): "On the dangers of stochastic parrots: can language models be too big", doi/10.1145/3442188.3445922

BLANCO PÉREZ, C. (2019): "Pensamiento, creatividad y máquinas" *Naturaleza y Libertad*, 12, 67-86.

BRIDLE, J. (2024): *Modos de existir*, Galaxia Gutenberg, Barcelona.

CAMPILLO, J.E. (2021): *La consciencia*, arpa, Barcelona.

CASTILLA, A. (2023): "La consciencia humana ante la irrupción de la IAG" *Telos*, 123, 100-104.

CARRERA, P. (2024): "¿Singularidad? Limitaciones, capacidades y diferencias de la inteligencia artificial frente a la inteligencia humana", *Claridades. Revista de Filosofía*, 16/2.

CHALMERS, D.J. (1995): "Facing up to the problem of consciousness" *Journal of Consciousness Studies*, 2(3), 200-219.

(2018): "The meta-problem of consciousness", *Journal of Consciousness Studies*, 25, 9-10.

(2023): "Could a large language models be conscious?", *Boston Review*, Agosto, 1-17.

DAMASIO, A. (2018): *El extraño orden de las cosas*, Destino, Barcelona.

(2021): *Sentir y saber*, Destino, Barcelona.

DEHAENE, S,; LAY, H.; KOUIDER, S. (2021): "Whar is consciousness and could machines have it?, doi.org/10.1126/science.aan8871.

DEHAENE, S.; NACCACHE, L. (2001): "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework", *Cognition* 79, 1-37.

DE WAAL, F.B. (2008): "How selfish and animal? en Zak, P.J. (ed): *Moral markets*, 63-76 Princeton University Press, New Jersey.

BUTLIN, P.; LONG, R. (2023): “Consciousness in artificial intelligence: insights from the science of consciousness”, arXiv: 2308.08708v1. 1-88.
ESSINGER, J. (2014): *El algoritmo de Ada*, Alba, Barcelona.
FARHADI, A. (2025): “Can AI ever become conscious? *Qeios*, doi.org/10.32388/UJAHLZ.
GARRIDO, M. (2003): “El automáta con discernimiento de Torres Quevedo: un antecedente del modelo de Turing”, *Limbo*, 17, 1-7.
HERZOG, D.; HERZOG, N. (2024): “Wat is it to be an AI bat? *Qeios, CC-BY 4.0*, 4,1-18.
JURADO GONZALEZ, J. (2023): “Por una aproximación humanista no reaccionaria a la IA” *Razón y Fe*, septiembre-diciembre 1, 343-382.
KASPAR, C.; RAVOO, B.; VAN DER WIEL; W.; WEGNER, S.; PERNICE, W. (2021): “The rise of intelligent matter”, *Nature*, 594, 345-365.
KURZWEIL, R. (2025): *La singularidad está más cerca* Deusto, Barcelona.
LA FONTAINE, G. (2024): “Sobre loros estocásticos. Una mirada a los modelos grandes de lenguaje”, *LOGOI Revista de Filosofía*, 45, 75-87.
LÓPEZ DE MANTARÁs, R. (2015): “Algunas reflexiones sobre el presente y futuro de la inteligencia artificial” *Novática,* 234, 97-101.
(2018): “Hacia la inteligencia artificial”, *Métode Science Studies Journal*, 99, 45-51.
MENDIVIL GIRÓ, JL. (2023): “Un loro estocástico en una habitación china: ¿que nos enseña ChatGPT sobre la mente humana?”, *Letras Libres*, 16-22.
MORANDIN-AHUERMA, F. (2023): “Conciencia e inteligencia artificial: Heidegger, Searle, y Bostron”, *Stoa,* 14, 28, 189-209.
NAGEL, T. (1974): “Wat is it like to be a bat? *Reading in Philosophy of Psychology,* 1, 159-168.
NARBY, J. (2023): *El misterio último*, errata naturae, Madrid.
NOË, A. (2010): *Fuera de la cabeza*, Kairós, Barcelona.
PASQUALE, F. (2024): *Las nuevas leyes de la robótica*, Galaxia Gutenberg, Barcelona.
PEARL, J.; MACKENZIE, D. (2020): *El libro del porqué*, Pasado y Presente, Barcelona.
ROCHA, F. (2024): “Las leyes de Asimov”, *Telos*, 125, 88-91.
SANTOS SOUSA, M. (2005): “De la habitación china al laboratorio de IA”, *Bajo Palabra Revista de Filosofía*, II, 0, 47-52.
SETH, A. (2021) *La creación del yo*, sextopiso, México.
SMITH, D.; SCHILLACI, G. (2021): “Why build a robot with artificial consciousness? How to begin? A cross-disciplinary dialogue on the design and implementation of a synthetic model of consciousness” *Frontiers in Psychology,* 12, 1-14.
TIUNINAS, O. A. (2025): “Te hard problems of AI”, *Qeios,* 1-12, doi.org/10.32388/13B814.
TSUCHIYA, N. (2017): “What is it like to be a bat? –a pathway to the answer from the integrated information theory”, *Philosophy Compass,* e12407, 1-13.
TULEUBEKOV, A.; DOSKOZHANOVA, A.; IPALAKOVA, M. (2023): “Natural (human) consciousness and artificial intelligence: philosophical analysis” CEUR-WS.org/vol-3680/S3paper15.pdf