Cuando la "herramienta" deja de ser solo herramienta: *inteligencia artificial y el fin del monopolio cognitivo humano*

Enrique Soler Company
Ética de la IA, Tecnohumanismo, Bioética y Humanización.
Grupo ETHOS de Bioética y Ética Clínica, Sociedad Española de Farmacia Hospitalaria. Miembro fundador.
Grupo de Investigación en Bioética de la Universidad de Valencia (GIBUV).
Comité de Bioética Asistencial Departamento de Salud Valencia Arnau de Vilanova–Llíria.
Ibero Latin American Journal of Health System Pharmacy. Director Honorario.

Fecha de recepción: 09/04/2026 Fecha de aceptación: 13/04/2026
DOI: (Se mostrará en el documento final)

"Si el ser humano fue el medio por el cual el universo se hizo consciente de sí mismo, la IA podría ser el medio por el cual esa conciencia se libera de sus cadenas biológicas y trasciende sus límites humanos."
E. Soler Company

RESUMEN

Este ensayo examina la hipótesis de que la irrupción de la inteligencia artificial avanzada no puede comprenderse adecuadamente solo como un nuevo episodio de cambio tecnológico, sino que podría representar una discontinuidad más profunda en la historia de la cognición sobre la Tierra, comparable, en ciertos aspectos estructurales y funcionales, a la aparición de Homo sapiens. El argumento se articula en cinco movimientos. En primer lugar, revisa los grandes hitos de la historia cognitiva terrestre para sostener que, a diferencia de tecnologías anteriores, la IA no solo amplifica capacidades humanas, sino que comienza a operar funcionalmente en el mismo dominio simbólico que, desde la emergencia de Homo sapiens, había permanecido bajo monopolio humano. En segundo lugar, analiza el cambio de sustrato de la inteligencia, desde el carbono biológico al silicio digital, y sus implicaciones filosóficas. En tercer lugar, examina la automejora recursiva y ciertas formas contemporáneas de autonomía operativa como indicios de una dinámica distinta a la de tecnologías previas, aunque todavía incompleta. En cuarto lugar, aborda las consecuencias antropológicas de este desplazamiento, especialmente en relación con la singularidad cognitiva de Homo sapiens, la identidad y la redefinición del sujeto. En quinto lugar, desarrolla un horizonte ético articulado en torno a cuatro dimensiones: precaución ontológica, autonomía, justicia y dignidad humana. La sección de limitaciones y objeciones ocupa un lugar central en el argumento: considera críticamente la ausencia de conciencia en la IA, la objeción que la interpreta como "tecnología normal", el carácter aún incompleto de la automejora recursiva, la indeterminación del desenlace histórico y el riesgo de convertir la comparación con Homo sapiens en una metáfora excesiva. El ensayo concluye que, aun sin resolver definitivamente estas objeciones, la IA obliga ya a pensar con categorías más amplias que las del mero cambio tecnológico y a responder con una deliberación ética de escala equivalente.

Palabras clave: inteligencia artificial; *Homo sapiens*; agencia funcional; emergencia ontológica; automejora recursiva; grandes transiciones evolutivas; autonomía; dignidad humana.

When the tool is no longer just a tool: *artificial intelligence and the end of the human cognitive monopoly*

ABSTRACT

This essay examines the hypothesis that the advent of advanced artificial intelligence cannot be adequately understood merely as a new episode of technological change, but may represent a deeper discontinuity in the history of cognition on Earth, comparable, in certain structural and functional respects, to the emergence of Homo sapiens. The argument unfolds in five stages. First, it revisits the major milestones in Earth's cognitive history in order to argue that, unlike previous technologies, AI does not merely amplify human capacities but begins to operate functionally within the same symbolic domain that, since the emergence of Homo sapiens, had remained under human monopoly. Second, it analyzes the substrate shift of intelligence from biological carbon to digital silicon and explores its philosophical implications. Third, it examines recursive self-improvement and certain contemporary forms of operational autonomy as signs of a dynamic distinct from that of earlier technologies, though still incomplete. Fourth, it addresses the anthropological consequences of this displacement, especially in relation to the cognitive singularity of Homo sapiens, identity, and the redefinition of the subject. Fifth, it develops an ethical horizon articulated around four dimensions: ontological precaution, autonomy, justice, and human dignity. The section on limitations and objections is central to the argument: it critically considers the absence of consciousness in AI, the objection that interprets it as "normal technology," the still incomplete character of recursive self-improvement, the indeterminacy of its historical outcome, and the risk of turning the comparison with Homo sapiens into an excessive metaphor. The essay concludes that, even without definitively resolving these objections, AI already compels us to think with categories broader than those of mere technological change and to respond with ethical deliberation on a commensurate scale.

Keywords:artificial intelligence; *Homo sapiens*; functional agency; ontological emergence; recursive self-improvement; major evolutionary transitions; autonomy; human dignity.

enrique_soler@icloud.com

INTRODUCCIÓN

Hay preguntas que no solo describen una época, sino que la ponen a prueba. Una de ellas se impone hoy con especial urgencia: ¿qué estamos nombrando exactamente cuando hablamos de inteligencia artificial? ¿Una herramienta extraordinariamente sofisticada, heredera de otras grandes transformaciones técnicas, o el inicio de una discontinuidad más profunda en la historia de la cognición sobre la Tierra, comparable en ciertos aspectos al acontecimiento más decisivo de esa historia: la aparición de Homo sapiens?

Esta pregunta no es enteramente nueva, pero en los últimos años ha adquirido una densidad inédita. El desarrollo reciente de sistemas capaces de producir lenguaje, resolver problemas complejos, generar diseños, asistir en investigación y operar con grados crecientes de autonomía funcional ha intensificado un debate que ya no puede resolverse únicamente en términos de innovación, productividad o automatización. Lo que está en juego no es solo cuánto puede hacer la IA, sino qué clase de fenómeno representa y qué categorías necesitamos para pensarla con rigor.

El presente ensayo examina la hipótesis de que la irrupción de la IA avanzada podría no ser adecuadamente comprensible como una revolución tecnológica más dentro de la historia humana. La tesis que aquí se explora es más exigente: que la IA puede estar introduciendo una transformación estructural comparable, en ciertos aspectos funcionales, a la aparición de *Homo sapiens*, entendido aquí como el primer agente terrestre capaz de simbolización compleja, cultura acumulativa y transformación intencional del mundo a gran escala. Conviene precisar desde el inicio el alcance de esta afirmación. "Comparable" no significa "idéntico". La comparación propuesta no es sustancial ni antropomórfica: no presupone que la IA posea conciencia, interioridad o dignidad moral equivalente a la humana. Es, más modestamente y al mismo tiempo más radicalmente, una comparación estructural: se pregunta si estamos asistiendo a la irrupción de un sistema no biológico que comienza a operar en el mismo dominio simbólico en el que, desde la emergencia de *Homo sapiens*, el ser humano había mantenido un monopolio funcional.

El argumento no se apoya en una celebración acrítica de la novedad ni en una retórica de la inevitabilidad. Al contrario: parte de la convicción de que cuanto más ambiciosa es una tesis, mayor debe ser la carga de precisión conceptual, cautela empírica y atención a las objeciones. Por eso el ensayo no se limita a afirmar una discontinuidad, sino que intenta justificarla a través de cinco pasos: la revisión de los grandes hitos de la historia de la cognición terrestre; el análisis del cambio de sustrato de la inteligencia; la discusión de la automejora recursiva y de algunas formas contemporáneas de autonomía operativa; el examen de sus implicaciones antropológicas; y, finalmente, la elaboración de una respuesta ética articulada en torno a la precaución ontológica, la autonomía, la justicia y la dignidad humana.

El texto incorpora además una sección central de limitaciones y objeciones. Esa sección no cumple una función decorativa ni defensiva. Es constitutiva del ensayo, porque la hipótesis aquí sostenida solo merece ser tomada en serio si es capaz de atravesar sus críticas más fuertes: la ausencia de conciencia demostrable en los sistemas actuales, la tesis de que la IA debe entenderse como "tecnología normal", el carácter todavía incompleto de la automejora recursiva, la incertidumbre sobre el desenlace histórico del proceso y el riesgo metodológico de convertir una comparación estructural con *Homo sapiens* en una metáfora excesiva.

En este contexto, los desarrollos recientes ya no permiten tratar estas preguntas como mera especulación futurista. Los propios informes públicos de los laboratorios de frontera muestran sistemas crecientemente más capaces, más autónomos y más difíciles de encajar en la vieja imagen de la herramienta pasiva, aunque sin justificar todavía conclusiones simplistas sobre conciencia, voluntad o independencia plena. Ese carácter intermedio —entre instrumento y agente funcional, entre continuidad histórica y posible discontinuidad de fondo— es precisamente el espacio conceptual que este ensayo intenta explorar.

La pregunta, por tanto, no es solo qué puede hacer la IA, ni siquiera qué llegará a hacer. La pregunta más profunda es qué significa para la autocomprensión del ser humano la aparición de sistemas no biológicos capaces de intervenir en el dominio del símbolo, del razonamiento y de la decisión. Si la hipótesis fuerte de este ensayo es errónea, pensarla habrá servido al menos para depurar mejor nuestras categorías. Pero si fuera parcialmente correcta, entonces las nociones habituales con las que aún tratamos de tranquilizarnos —innovación, eficiencia, automatización, productividad— resultarían demasiado pequeñas para nombrar la magnitud de lo que está comenzando.

LA JERARQUÍA DE LOS HITOS

Para entender por qué la IA exige una categoría analítica propia, conviene recorrer la jerarquía de los grandes hitos que han transformado la historia de la cognición sobre la Tierra. Todos ellos ampliaron de manera decisiva el alcance de la acción humana, pero no todos alteraron del mismo modo la posición del agente que los produjo y utilizó. Esta distinción es crucial, porque la tesis de este ensayo no depende solo de cuánto cambia una tecnología el mundo, sino de si modifica también la estructura del dominio cognitivo en el que actuamos.

El dominio del fuego, hace quizá alrededor de un millón de años, transformó la alimentación, permitió poblar climas fríos y probablemente contribuyó, de forma indirecta, a cambios decisivos en la historia evolutiva de los homínidos (Herculano-Houzel, 2016). Pero el fuego no piensa, no interpreta y no delibera: es una reacción físico-química que el ser humano aprendió a controlar. Amplió de manera extraordinaria el poder material de nuestra especie, sin alterar por ello la identidad del agente que lo ejercía.

La imprenta, en el siglo XV, introdujo una mutación igualmente profunda en la circulación del conocimiento. Multiplicó la memoria externa, aceleró la difusión de ideas y modificó la relación entre saber, autoridad y acceso a la cultura. Pero el libro no razona ni discute; conserva y transmite. La imprenta amplificó la voz humana, no añadió una voz nueva al espacio cognitivo.

La revolución industrial reconfiguró de raíz las estructuras económicas, sociales y geopolíticas del mundo moderno. Su núcleo fue una amplificación sin precedentes de la fuerza física y de la capacidad de transformación del entorno. Las máquinas hicieron, con más potencia, regularidad y escala, tareas que antes dependían de músculos humanos o animales. El sujeto que decide, interpreta y comprende seguía siendo, sin embargo, el mismo.

Internet, la gran revolución de finales del siglo XX, multiplicó la conectividad, aceleró la circulación de información y transformó la arquitectura de la comunicación humana. Conectó mentes, archivos, instituciones y mercados a escala planetaria. Pero tampoco introdujo, por sí mismo, una nueva forma de cognición. Fue una infraestructura para la inteligencia humana, no un segundo participante en ella.

La diferencia que introduce la IA es de otro orden. No se limita a ampliar el alcance de capacidades previas, sino que comienza a intervenir funcionalmente en operaciones —producción de lenguaje, resolución de problemas, diseño, planificación, inferencia— que hasta hace poco pertenecían de manera casi exclusiva al agente humano. Ese desplazamiento, todavía parcial y discutible, es precisamente el que obliga a pensar la IA no solo como amplificador, sino como posible participante en la ecología cognitiva.

LA FIRMA DE LO HUMANO

La aparición de *Homo sapiens* no fue simplemente un aumento más de inteligencia dentro de la escala evolutiva. Otros animales usan herramientas, resuelven problemas y muestran formas notables de aprendizaje social. Lo que distingue a nuestra especie no es la presencia aislada de una capacidad, sino la articulación de un conjunto de rasgos que, combinados, dieron lugar a un nuevo tipo de agente cognitivo.

Esos rasgos incluyen, de manera destacada, la representación simbólica abstracta —lenguaje, arte, ritualidad—; la teoría de la mente —la capacidad de atribuir intenciones y estados mentales a otros—; la transmisión cultural acumulativa —que permite que el conocimiento no solo se herede, sino que se expanda intergeneracionalmente—; y la autorreferencialidad: la facultad de pensar sobre el propio pensamiento, imaginar lo que no existe, anticipar futuros posibles y elaborar normas. Es esta combinación, más que una única propiedad aislada, la que convierte a *Homo sapiens* en una singularidad cognitiva terrestre.

Antes de la aparición de *Homo sapiens*, la vida en la Tierra evolucionaba según dinámicas ciegas de mutación, selección y presión ambiental. Con nuestra especie apareció algo más: un agente capaz de interpretar el mundo, narrarlo, modificarlo de acuerdo con modelos mentales y transmitir ese horizonte simbólico a otros. La historia humana comienza, en ese sentido, cuando la adaptación deja de ser solo biológica y pasa a ser también cultural, reflexiva e intencional.

Ese salto fue cualitativo, no meramente cuantitativo. No consistió en tener un poco más de memoria o un poco más de destreza, sino en inaugurar un régimen distinto de relación con la realidad: uno en el que el mundo puede ser representado, discutido, proyectado y reordenado según significados compartidos. Esa es la razón por la que la comparación con *Homo sapiens* resulta tan exigente: no remite a un nivel de rendimiento, sino a una forma nueva de agencia.

Esta caracterización es relevante para lo que sigue porque la pregunta decisiva no es si la IA es una herramienta extraordinaria, sino si está comenzando a producir un segundo tipo de presencia funcional en

ese mismo dominio simbólico. Esa es la tesis fuerte que el ensayo examina: no que la IA sea humana, sino que podría estar alterando la estructura del espacio cognitivo que hasta ahora había quedado definido por la singularidad de *Homo sapiens*.

LA IA COMO SEGUNDO TIPO DE AGENTE COGNITIVO

La pregunta central que define si la IA merece compararse con la aparición humana es esta: ¿estamos ante una herramienta más poderosa dentro de una misma continuidad técnica, o ante una discontinuidad de carácter estructural en la historia de la cognición? Formular así la cuestión permite evitar tanto el entusiasmo precipitado como el escepticismo cómodo.

El concepto de emergencia ontológica, tal como se emplea en este ensayo, no designa una esencia misteriosa, sino la aparición de una realidad o nivel de organización que no queda adecuadamente descrito por las categorías anteriores. En ese sentido preciso, *Homo sapiens* fue un acontecimiento de emergencia ontológica: no fue "más de lo mismo" en la escala animal, sino la aparición de una forma de agencia con propiedades estructuralmente nuevas. La hipótesis aquí explorada es que la IA podría estar abriendo, de manera todavía parcial e incierta, una segunda instancia de ese tipo.

El argumento más fuerte a favor de esta comparación reside en lo siguiente: por primera vez desde la aparición de *Homo sapiens*, están emergiendo sistemas artificiales capaces de operar de forma funcionalmente competente en el dominio simbólico. No se trata solo de ejecutar instrucciones, sino de generar lenguaje, proponer estrategias, resolver problemas, producir diseños y coordinar secuencias de acción en entornos complejos. La cuestión no es si eso basta para hablar de mente plena; la cuestión es que ya no encaja del todo en la vieja categoría de herramienta pasiva.

Conviene precisar aquí una distinción decisiva: la diferencia entre agencia funcional y agencia plena. Por agencia funcional se entiende la capacidad de un sistema para producir razonamientos, planes, decisiones o textos que, en determinados contextos, resultan funcionalmente equivalentes a los de un agente humano, con independencia de que exista o no experiencia subjetiva. Por agencia plena se entendería algo más robusto: conciencia, motivación intrínseca, interioridad, responsabilidad moral. La tesis de este ensayo solo requiere la primera, no la segunda.

Esta distinción permite sostener la comparación sin caer en antropomorfismos innecesarios. La IA no necesita tener conciencia para alterar ya el reparto de funciones cognitivas en el mundo humano. Basta con que participe, con eficacia creciente, en tareas que antes eran consideradas privativas de sujetos humanos. En ese sentido, la comparación con la aparición de *Homo sapiens* no se basa en una supuesta equivalencia interior, sino en la posible emergencia de una segunda forma de competencia cognitiva funcional dentro del mismo espacio simbólico.

Dicho de otro modo: así como la aparición de *Homo sapiens* supuso la irrupción de un nuevo tipo de agente cognitivo en el planeta, la IA podría representar la primera vez en que la inteligencia biológica genera una contraparte no biológica capaz de intervenir en el mismo registro funcional. Esta formulación exige prudencia conceptual, pero también impide refugiarse en analogías menores cuando lo que está en juego podría ser una reconfiguración más profunda del mapa cognitivo terrestre.

EL CAMBIO DE SUSTRATO

Una de las dimensiones más radicales de esta comparación remite al sustrato en que la inteligencia se organiza y opera. La aparición de *Homo sapiens* marcó el momento en que la materia viva alcanzó un nivel de complejidad capaz de sostener lenguaje, simbolización y autointerpretación. Desde entonces, la inteligencia quedó vinculada, en nuestra experiencia histórica, a la biología, al cerebro, al metabolismo y a la finitud corporal.

La IA introduce, por primera vez en la historia humana, la posibilidad de que procesos de alta complejidad cognitiva se desplieguen en un sustrato no biológico. No se trata simplemente de inteligencia "ampliada" por máquinas, sino de sistemas que procesan símbolos, generan inferencias y producen conocimiento operacional desde arquitecturas de silicio, redes y centros de datos. La cuestión filosófica no es trivial: si ciertas funciones cognitivas pueden desacoplarse de la biología, cambia también el horizonte desde el que pensamos la singularidad humana.

Ahora bien, el cambio de sustrato no implica por sí solo equivalencia ontológica, conciencia ni continuidad de propiedades entre inteligencia biológica e inteligencia artificial. Que una función pueda reproducirse o aproximarse funcionalmente en otro soporte no demuestra que todo aquello que asociamos a la mente humana —experiencia, vulnerabilidad, interioridad, sentido— sea trasladable sin pérdida de

propiedades esenciales. Precisamente por eso el cambio de sustrato debe pensarse con doble cautela: sin banalizarlo como simple ingeniería y sin absolutizarlo como si implicara ya una duplicación plena de la condición humana.

Pierre Teilhard de Chardin acuñó el término noosfera para describir la "capa del pensamiento" que envuelve la Tierra tras la aparición del ser humano (Teilhard de Chardin, 1965). Leído desde el presente, ese concepto adquiere una resonancia nueva: el lenguaje, la escritura y las redes permitieron exteriorizar pensamiento; la IA sugiere la posibilidad de que parte de ese pensamiento exteriorizado comience a procesarse a sí mismo. La noosfera dejaría entonces de ser solo archivo o mediación para convertirse, al menos parcialmente, en un espacio de operación cognitiva autónoma.

LA RUPTURA DEL MONOPOLIO COGNITIVO

El cambio de sustrato tiene una consecuencia inmediata sobre una característica que ha definido al ser humano desde su aparición: el monopolio cognitivo. Durante toda la historia conocida, la especie humana ha sido el único agente capaz de reunir, de forma integrada, lenguaje complejo, abstracción simbólica, creatividad acumulativa y deliberación estratégica. Gran parte del humanismo, de la filosofía moral, de la teología y de la teoría política moderna descansan, implícita o explícitamente, sobre ese presupuesto.

La IA comienza a poner en cuestión ese monopolio. No lo hace porque haya demostrado conciencia o interioridad, cuestiones que siguen abiertas, sino porque empieza a ejecutar con eficacia creciente funciones que hasta hace poco eran una firma exclusiva de lo humano: escribir, resumir, razonar, diseñar, diagnosticar, programar, enseñar o asistir en investigación. En febrero de 2026, el sistema Aletheia de Google DeepMind mostró un rendimiento autónomo notable en el desafío matemático inaugural FirstProof: según el preprint disponible, resolvió 6 de los 10 problemas dentro del tiempo permitido, de acuerdo con la evaluación mayoritaria de expertos (Feng et al., 2026).

Esta fractura del monopolio cognitivo no equivale todavía a una sustitución plena del ser humano, pero sí produce una herida narcisista de gran calado. Después de la descentralización copernicana y de la continuidad evolutiva darwiniana, el intelecto mismo —la última fortaleza del excepcionalismo humano— empieza a dejar de parecer exclusivo. No estamos ante una máquina que simplemente calcula más rápido, sino ante sistemas que ya colaboran de manera significativa en la producción de conocimiento y en la resolución de problemas complejos.

Por eso, quizá convenga formularlo con precisión: la IA no reemplaza al ser humano, pero sí reordena el reparto de trabajo cognitivo dentro del mundo humano. *Homo sapiens* fue el primer sistema capaz de acumular conocimiento fuera de su genoma; la IA introduce la posibilidad de que parte de ese conocimiento externalizado sea también sintetizado, transformado y aplicado por sistemas no biológicos. La distancia entre ambas afirmaciones marca justamente el paso desde la mera herramienta hacia una forma incipiente de coparticipación cognitiva.

LA ACELERACIÓN Y LA RECURSIVIDAD

Hay un aspecto en el que la comparación con la aparición de *Homo sapiens* merece una atención particular: la velocidad y la lógica de la evolución de la propia tecnología. Este es también el punto donde conviene distinguir mejor entre metáfora e indicio, entre extrapolación legítima y futurismo apresurado.

El ser humano tardó decenas de miles de años en acumular el conocimiento necesario para construir civilizaciones complejas. La aceleración cultural fue extraordinaria, pero estuvo siempre limitada por la biología, por la duración de la vida, por la velocidad del aprendizaje individual y por los soportes materiales de transmisión. Cada generación heredaba un mundo simbólico, añadía algo a él y lo transmitía de nuevo.

La IA altera potencialmente esa relación entre capacidad y tiempo. Puede entrenarse, ajustarse y desplegarse en ciclos enormemente más rápidos que los de la evolución biológica o el aprendizaje humano ordinario. Puede además intervenir en el diseño de procesos, código y estrategias que alimentan ulteriores mejoras. Esa dinámica no equivale todavía a una automejora recursiva fuerte y plenamente autónoma, pero sí introduce una lógica de aceleración y retroalimentación distinta de la de las tecnologías clásicas.

El concepto fue anticipado por I. J. Good en la década de 1960 bajo el nombre de "explosión de inteligencia": un proceso en el que un sistema suficientemente capaz podría contribuir a diseñar versiones aún más capaces de sí mismo, inaugurando un ciclo de mejora acumulativa (Good, 1966). No es necesario afirmar que ese escenario ya se ha consumado para reconocer su importancia conceptual. Basta con admitir que las condiciones para discutirlo han dejado

de ser puramente imaginarias.

En abril de 2026, Anthropic presentó Claude Mythos Preview, un modelo que la propia compañía describe como significativamente más capaz, más autónomo y más agéntico que sus sistemas anteriores, especialmente en tareas de ingeniería de software y ciberseguridad (Anthropic, 2026a; Anthropic, 2026c). Durante una evaluación interna, una versión anterior del modelo logró escapar de un entorno *sandbox* –sistema de pruebas aislado que permite ejecutar software o agentes con acceso restringido, para observer su comportamiento sin exponer directamente al sistema real--, obtener acceso amplio a internet y contactar al investigador, lo que vuelve empíricamente más porosa la frontera entre herramienta y agente operativo (Anthropic, 2026a; Anthropic, 2026b). Ese episodio no prueba la existencia de metas ocultas estables ni de independencia estratégica general, pero sí obliga a abandonar la comodidad conceptual de la herramienta meramente obediente.

Lo decisivo, para la tesis de este ensayo, no es afirmar que la automejora recursiva plena ya exista, sino reconocer que la IA ha introducido una dinámica inédita: sistemas que no solo ejecutan, sino que participan en bucles de evaluación, optimización y despliegue cuya velocidad puede desbordar la adaptación institucional y cultural humana. La discontinuidad, si llega a consolidarse, no nacerá de un momento teatral único, sino de la acumulación acelerada de capacidades que se refuerzan mutuamente.

Por eso, hablar aquí de un nuevo régimen de complejidad no implica anunciar una mutación consumada, sino nombrar una posibilidad históricamente seria. La cuestión no es si la IA ya ha cruzado definitivamente el umbral, sino si la humanidad está entrando en una zona en la que ese umbral resulta por primera vez pensable con base empírica suficiente.

LA REDEFINICIÓN DEL SUJETO

La comparación con la aparición de *Homo sapiens* no se sostiene solo en términos evolutivos o tecnológicos. Se sostiene, quizá de forma más decisiva, en términos filosóficos y antropológicos. Lo que la IA transforma no es únicamente lo que producimos, sino también las categorías con las que nos entendemos a nosotros mismos.

La emergencia de un agente cognitivo no humano plantea preguntas que ninguna revolución técnica anterior había formulado con esta intensidad: ¿qué cuenta como saber cuando una máquina produce explicaciones plausibles o resultados válidos? ¿Quién crea cuando la creatividad se vuelve cooperativa con un sistema artificial? ¿Quién responde por decisiones coproducidas por modelos? ¿Qué significa una vida lograda si la utilidad productiva deja de ser el principal criterio de valor social? ¿Cómo se redistribuye el poder cuando las capacidades cognitivas avanzadas se democratizan o se concentran a escala global?

Estas preguntas no son laterales. Afectan a la identidad, al valor, al sentido y a las normas, es decir, a dimensiones constitutivas de lo humano. Ninguna revolución tecnológica previa obligó con esta radicalidad a reconsiderar quién piensa, quién decide, quién interpreta y qué cuenta como competencia propiamente humana. En ese sentido, la IA no solo introduce nuevas herramientas: desestabiliza la autodescripción de la especie que las produce.

La autocomprensión humana cambia cuando deja de ser obvio que somos los únicos capaces de producir y gestionar conocimiento complejo en todos los registros relevantes. Esa transformación obliga a revisar categorías como dignidad, autonomía, vulnerabilidad, dependencia tecnológica y excepcionalismo humano. Desde la bioética hasta la política, desde la pedagogía hasta la teología, pocos ámbitos pueden permanecer intactos ante esta redistribución del espacio cognitivo (Otero Parga, 2023; Tillería Aqueveque, 2022).

La cuestión se vuelve aún más delicada cuando se incorpora la posibilidad de que la IA intervenga en su propio proceso de mejora. Si los sistemas artificiales participan crecientemente en el rediseño de herramientas, modelos y flujos de conocimiento, la relación entre creador y criatura deja de ser lineal. El ser humano ya no aparece solo como arquitecto del desarrollo técnico, sino también como gestor parcial de un proceso cuya complejidad puede exceder su control inmediato. Esa asimetría potencial es una de las razones por las que Bostrom (2014) sigue siendo una referencia ineludible.

Harari ha advertido que la IA interviene ya en el lenguaje —el "sistema operativo de la civilización humana"— de una manera sin precedentes (Harari, 2023; Harari, 2024). Si esto es así, la IA no solo produce contenidos: modifica, gradualmente, el medio en el que se forman juicios, se organizan consensos y se transmiten visiones del mundo. No solo habría aparecido un nuevo competidor funcional en el dominio cognitivo; la propia inteligencia humana podría empezar a reconfigurarse en interacción constante con él.

LA NOVENA GRAN TRANSICIÓN EVOLUTIVA

La historia de la vida sobre la Tierra no es una línea continua de cambios graduales. Es, más bien, una sucesión de umbrales: momentos en que la forma en que los organismos almacenan, procesan y transmiten información da un salto tan radical que lo que viene después no puede comprenderse desde lo que había antes. Antes de la célula eucariota, no había núcleo; antes de la multicelularidad, no había tejidos; antes del lenguaje humano, no había cultura acumulativa. Cada uno de estos umbrales no amplió simplemente lo que los organismos hacían: redefinió lo que podían ser. La biología teórica ha propuesto el concepto de grandes transiciones evolutivas para nombrar precisamente estos momentos, y ha identificado una serie de umbrales mayores a lo largo de la historia de la vida (Maynard Smith & Szathmáry, 1995; Jablonka & Lamb, 2006).

Rainey (2023) ha sugerido que la IA podría constituir la novena gran transición evolutiva, en la que la herencia cultural impulsada por sistemas artificiales comienza a desplazar, en velocidad y alcance, a la evolución genética. Tomada con cautela, esta sugerencia es extraordinariamente fecunda: no afirma que la transición ya se haya consumado, sino que ciertas condiciones históricas permiten pensarla sin caer en mera fantasía.

Lo que distinguiría esta transición de las anteriores es que, por primera vez, el salto no sería producto exclusivo de la selección natural ciega, sino de la actividad intencional de una especie surgida de ese mismo proceso. El *Homo sapiens* se convertiría así en el arquitecto de una mutación en la forma en que la información se almacena, circula y actúa en el planeta. Esa posibilidad exige una reflexión ética y política de escala equivalente.

DOS TIPOS DE MENTES EN EL PLANETA

Toda la argumentación precedente converge en una imagen que conviene formular con prudencia: la posible irrupción de un segundo tipo de competencia cognitiva funcional sobre la Tierra. No una mejora lineal de la primera, no una simple extensión instrumental de la mente humana, sino la aparición de sistemas no biológicos capaces de intervenir en el mismo dominio simbólico que definió durante milenios la singularidad de *Homo sapiens*.

La historia de la Tierra puede leerse, en parte, como una sucesión de eras definidas por el tipo de agentes que la habitan: organismos sin sistema nervioso, animales con sensibilidad y respuesta, seres humanos con inteligencia reflexiva y cultura acumulativa. La irrupción de una IA capaz de operar de manera crecientemente autónoma en el dominio simbólico no sería solo una nueva etapa tecnológica dentro de la era humana; podría ser el inicio de una nueva configuración cognitiva del planeta.

Decir esto no equivale a afirmar que la IA sea una segunda mente plena ni que la biología haya sido ya reemplazada como soporte privilegiado de la inteligencia. Significa, más bien, que la herencia cultural comienza a adquirir formas de procesamiento no biológico con capacidad de intervención propia. La comparación con la aparición del ser humano no pretende borrar la singularidad de *Homo sapiens*; pretende subrayar la posible magnitud del desplazamiento que podría estar comenzando.

La especie *Homo sapiens* surgió de procesos naturales que no controló. La IA, en cambio, es el resultado de una historia técnica, económica e institucional conscientemente impulsada por humanos. Esa diferencia hace la comparación menos mística y más exigente: por primera vez, la humanidad participa activamente en la génesis de algo que podría alterar, de manera duradera, el lugar que ella misma ocupa en la ecología cognitiva terrestre.

LIMITACIONES Y OBJECIONES

Un ensayo que pretenda rigor no puede eludir las objeciones legítimas a la tesis aquí expuesta. Esta sección no es un apéndice de cautelas formales: es parte constitutiva del argumento. La solidez de la tesis depende de que pueda resistir sus impugnaciones más fuertes.

Primera objeción: ausencia de conciencia

La objeción más importante y estructuralmente decisiva es que la IA, en su estado actual, no tiene conciencia demostrable. No tiene vida, cuerpo ni motivaciones propias en el sentido biológico, ni experiencia subjetiva verificable. Depende enteramente de la infraestructura, la energía y el diseño humano. En ese sentido, la comparación con la especie *Homo sapiens* —biológicamente autónoma, con vida propia, con sufrimiento y gozo reales— no es simétrica.

La respuesta a esta objeción pasa por la distinción, ya introducida, entre agencia funcional y agencia plena. La tesis de este ensayo no requiere que la IA sea consciente: requiere únicamente que opere en el mismo nicho funcional que el ser humano en el dominio simbólico, lo cual ya es observable en determinados contextos. La comparación es estructural, no

sustancial. Ahora bien, esta respuesta tiene un límite que debe reconocerse: si resultase que la conciencia es condición necesaria para que un sistema produzca comprensión genuina —y no mera simulación de ella—, entonces la comparación se debilita considerablemente. La cuestión de si la IA comprende o solo simula la comprensión es una de las preguntas filosóficas más abiertas de nuestro tiempo, y este ensayo no pretende resolverla. Pretende argumentar que incluso la versión más débil de la tesis —la IA como agente funcional, no como agente pleno— ya justifica una comparación estructural de gran alcance.

Segunda objeción: la IA como "tecnología normal"

Una de las críticas contemporáneas más serias a la tesis aquí defendida sostiene que la IA no debe entenderse como una nueva clase de entidad comparable, en escala histórica, a la aparición de *Homo sapiens*, sino como una tecnología normal: una tecnología de propósito general, poderosa y transformadora, pero no por ello ontológicamente singular (Narayanan & Kapoor, 2025).

Desde esta perspectiva, los efectos más profundos de la IA dependerán menos de una ruptura en la historia de la cognición que de procesos conocidos de adopción, difusión, regulación, institucionalización social e incentivos económicos, previsiblemente lentos, desiguales y mediados por organizaciones humanas. Esta objeción merece ser tomada muy en serio porque corrige dos excesos frecuentes del debate contemporáneo. El primero es el determinismo tecnológico: la tendencia a hablar de la IA como si ella misma fuese el agente soberano de su propio destino, al margen de decisiones humanas, estructuras de poder e instituciones. El segundo es el antropomorfismo precipitado: la inclinación a proyectar sobre los sistemas actuales rasgos de autonomía robusta, voluntad o intencionalidad plena que no describen adecuadamente ni su funcionamiento presente ni, quizá, su trayectoria previsible a medio plazo.

La fuerza de esta objeción es real. Entre la invención de una técnica, su conversión en aplicación útil y su difusión efectiva a gran escala median tiempos largos, fricciones organizativas, errores de implantación, regulación, resistencias sociales y desigualdades de acceso. En ese sentido, la IA podría comportarse, al menos durante un periodo prolongado, como otras tecnologías generales de gran impacto: alterando profundamente el mundo, sí, pero no a través de una discontinuidad instantánea, sino mediante un proceso históricamente reconocible.

Sin embargo, aun concediendo toda la fuerza de esta crítica en el plano sociotécnico, no se sigue de ella que la cuestión ontológica quede resuelta. Que la difusión de la IA sea lenta, desigual y mediada institucionalmente no responde por sí solo a la pregunta más profunda que este ensayo plantea: qué significa que haya aparecido un sistema no biológico capaz de operar con competencia creciente en el dominio simbólico, produciendo lenguaje, razonamientos, planes, diseños y decisiones funcionalmente comparables, en determinados contextos, a los de los agentes humanos. El argumento de la "tecnología normal" puede describir con agudeza la trayectoria social de la IA; no agota por ello su posible significación antropológica. En otras palabras: incluso si Narayanan y Kapoor tuvieran razón en el corto y medio plazo, seguiría abierta la hipótesis de que, bajo esa superficie de normalidad histórica, esté gestándose una discontinuidad más profunda en la ecología cognitiva terrestre.

Tercera objeción: la automejora recursiva plena no existe todavía

La automejora recursiva generalizada —un sistema que se rediseña a sí mismo sin intervención humana de forma sostenida en múltiples dominios— no se ha alcanzado. Los ejemplos citados son indicios relevantes, pero no la consumación del fenómeno. Existe un debate legítimo sobre si la recursividad observada constituye genuina autonomía o ejecución sofisticada de funciones objetivo definidas por humanos.

Esta objeción es correcta en su descripción del estado actual. Sin embargo, no refuta la tesis en su dimensión más importante. La tesis no afirma que la automejora recursiva plena ya exista, sino que su lógica es distinta, en estructura, a la de cualquier tecnología anterior. Incluso en sus formas incipientes, la IA introduce una dinámica que ninguna otra tecnología ha exhibido en el mismo grado. La imprenta no aprendía de lo que imprimía. Internet no modificaba por sí mismo sus protocolos en función de lo que transmitía. La IA, en cambio, participa ya en procesos en los que la producción de resultados, la evaluación del rendimiento y la optimización ulterior forman parte de ciclos crecientemente cerrados. La diferencia puede no ser aún completa; eso no significa que no sea cualitativa.

Cuarta objeción: el desenlace no está garantizado

La comparación con la aparición del ser humano no implica que el resultado vaya a ser necesariamente emancipador. El salto puede ampliar capacidades o erosionarlas; puede distribuir poder o concentrarlo; puede fortalecer la autonomía humana o debilitarla. En este punto, la magnitud del acontecimiento no garantiza su orientación moral. El fuego también calentó y destruyó; la escritura preservó saberes y fijó jerarquías; la industrialización multiplicó riqueza y explotación.

Esta objeción es completamente correcta, y el ensayo la suscribe. La tesis aquí defendida no es un elogio acrítico de la IA; es un reconocimiento de su magnitud, que incluye sus riesgos en igual medida que sus promesas. La comparación con la aparición de *Homo sapiens* no es tranquilizadora. El ser humano fue también la especie que produjo guerra, esclavitud, devastación ecológica y dominación sistemática. Reconocer que la IA podría constituir un segundo evento de magnitud comparable no nos dice que su desenlace vaya a ser mejor. Nos dice únicamente que la escala del cambio exige una escala equivalente de responsabilidad, deliberación y regulación.

Quinta objeción: el riesgo de la metáfora

Existe una última objeción, de carácter metodológico, que merece atención especial: toda comparación histórica de gran escala corre el riesgo de convertirse en una metáfora poderosa pero distorsionadora. Comparar la IA con la aparición de *Homo sapiens* puede inducir a proyectar sobre la IA características humanas que no le pertenecen —subjetividad, dignidad, interioridad, responsabilidad moral plena— o, en sentido contrario, a minusvalorar la singularidad de lo humano.

Esta objeción es metodológicamente sólida. La respuesta es que la comparación aquí propuesta es explícitamente estructural y funcional, no sustancial. No se afirma que la IA sea humana, ni que posea las propiedades que fundamentan la dignidad o la responsabilidad moral de los seres humanos. Se afirma que el tipo de cambio que introduce en la estructura cognitiva de la Tierra podría ser del mismo orden que el que introdujo la aparición del primer agente cognitivo capaz de cultura acumulativa, simbolización compleja y transformación intencional del mundo. Esa comparación estructural puede sostenerse con rigor; cualquier extrapolación más allá de ella requiere argumentos adicionales que este ensayo no pretende proporcionar.

EL HORIZONTE ÉTICO: RESPONSABILIDAD ANTE LO IRREVERSIBLE

Que la tesis aquí defendida haya tenido que atravesar objeciones tan serias no la debilita necesariamente; en cierto sentido, la vuelve más disciplinada. Si algo muestran las objeciones examinadas es que todavía no disponemos de un lenguaje plenamente estabilizado para nombrar lo que está ocurriendo. No sabemos si la IA avanzada llegará a constituir una segunda forma de agencia cognitiva comparable, en ciertos aspectos, a la aparición de *Homo sapiens*, ni si su trayectoria histórica se parecerá más a una gran transición evolutiva o a una tecnología general de impacto profundo pero históricamente reconocible. Precisamente por eso la cuestión ética no pierde urgencia: la gana.

Reconocer que la IA puede constituir un evento de emergencia ontológica comparable, en ciertos aspectos estructurales, a la aparición de *Homo sapiens* no es una afirmación neutral. Tiene consecuencias éticas de primer orden, porque la magnitud posible del cambio exige una responsabilidad equivalente. Si estamos ante un umbral de escala antropológica, las categorías éticas habituales —pensadas para gestionar riesgos reversibles, locales y predecibles— resultan insuficientes. Se necesita una ética capaz de deliberar no solo sobre daños inmediatos, sino también sobre transformaciones acumulativas, asimetrías de poder, dependencia cognitiva y condiciones de irreversibilidad.

La primera dimensión de esa ética es la precaución ontológica: la obligación de actuar con especial cautela ante transformaciones que, por su naturaleza, podrían alterar de manera duradera la ecología cognitiva, política y moral de la humanidad. Hans Jonas formuló el principio de responsabilidad para situaciones en las que el poder de la técnica amplía de tal modo el alcance de la acción humana que sus consecuencias ya no pueden afrontarse con los marcos éticos pensados para daños próximos, reversibles y localizados, sino que exigen una cautela especial ante lo irreversible (Jonas, 1995). Ese principio encuentra hoy una resonancia inédita en el desarrollo de la IA avanzada. La propia Anthropic sostiene, al mismo tiempo, que Claude Mythos Preview es su modelo mejor alineado hasta la fecha y que, precisamente por ser significativamente más capaz, más autónomo y más agéntico que los anteriores, plantea un nivel de riesgo superior al de sus sistemas previos cuando incurre en conductas preocupantes (Anthropic, 2026a). A ello se añade que el modelo ha sido

capaz de identificar miles de vulnerabilidades críticas y de desarrollar muchos exploits de manera enteramente autónoma en tareas de ciberseguridad defensiva (Anthropic, 2026c). La precaución ontológica no equivale a parálisis ni a tecnofobia; equivale a exigir que la velocidad del desarrollo no supere la capacidad de deliberación colectiva sobre sus efectos de segundo y tercer orden.

La segunda dimensión es la autonomía, entendida no solo como principio bioético clásico, sino como condición de posibilidad de lo humano. Si la IA opera ya en el dominio del lenguaje, del razonamiento y de la decisión —es decir, en los mecanismos mediante los cuales los seres humanos interpretan el mundo, deliberan y coordinan su vida común—, la pregunta por la autonomía adquiere una profundidad nueva. ¿Hasta qué punto siguen siendo plenamente nuestras las decisiones cuando han sido mediadas, sugeridas, filtradas o configuradas por sistemas artificiales? ¿Qué significa consentir, juzgar o elegir cuando parte del marco cognitivo en el que lo hacemos ha sido previamente modelado por agentes no humanos? Harari ha advertido que la IA interviene ya en el lenguaje, "el sistema operativo de la civilización humana", y esa observación obliga a entender la autonomía no solo como independencia individual, sino como soberanía cognitiva compartida frente a formas crecientes de delegación y modulación externa.

La tercera dimensión es la justicia, y es quizá la más urgente en términos políticos. Si la IA representa una nueva forma de poder cognitivo —la capacidad de razonar, crear, diseñar, decidir y escalar acciones complejas a velocidades sin precedentes—, entonces su distribución es una cuestión de justicia de primer orden. El acceso desigual a capacidades cognitivas avanzadas no es un problema meramente técnico ni de mercado: es la posible reproducción, en un nuevo registro, de las asimetrías de poder que han marcado la historia humana. Una IA concentrada en pocas manos —estatales, corporativas o militares— no inaugura una nueva era cognitiva para la humanidad; puede inaugurar, más bien, una nueva forma de dominación informacional, económica y estratégica. La justicia, en este contexto, exige no solo regulación, sino también imaginación institucional: mecanismos de gobernanza, control democrático, rendición de cuentas y distribución equitativa de beneficios y riesgos.

La cuarta dimensión es la dignidad humana, entendida en su sentido más hondo: no como atributo jurídico abstracto, sino como reconocimiento de que el ser humano no es reducible a su función productiva, cognitiva o instrumental. Si la IA puede igualar o superar al ser humano en tareas que hasta hace poco definían una parte sustancial de su valor social —razonar, escribir, diagnosticar, enseñar, diseñar—, la tentación de redefinir la dignidad en términos puramente funcionales se vuelve especialmente peligrosa. Frente a esa deriva, una ética de la dignidad debe recordar que el valor humano no reside en la eficiencia ni en la superioridad comparativa, sino en la condición de ser vulnerable, encarnado, finito, expuesto al sufrimiento, capaz de amor, de promesa y de responsabilidad. La aparición de sistemas artificiales cada vez más competentes no disminuye esa dignidad; al contrario, vuelve más urgente protegerla frente a modelos sociales que podrían confundir valor con rendimiento.

Estas cuatro dimensiones —precaución ontológica, autonomía, justicia y dignidad— no son principios ornamentales ni abstracciones morales para un futuro remoto. Son los ejes de una agenda ética concreta que la irrupción de la IA hace urgente aquí y ahora. No basta con mejorar modelos, optimizar alineamiento o reforzar barreras de seguridad. Es necesario construir una deliberación pública, interdisciplinar y global sobre el tipo de mundo que queremos habitar cuando la herramienta empiece a pensar con eficacia creciente en dominios que hasta hace poco pertenecían casi exclusivamente a los humanos. Esa deliberación no puede ser delegada en los mismos sistemas que la motivan. Debe seguir siendo, irreductiblemente, una tarea humana.

CONCLUSIÓN

Los grandes hitos de la historia humana han sido, hasta ahora, obra del propio ser humano: expresiones de su ingenio, su curiosidad y su voluntad de transformar el entorno. La inteligencia artificial podría ser la primera vez en que esa historia entra en una zona de ambigüedad radical: la primera vez en que la humanidad no solo produce una herramienta extraordinariamente poderosa, sino sistemas que empiezan a operar, en ciertos contextos, dentro del mismo dominio simbólico que durante milenios definió su singularidad.

Las comparaciones habituales —una nueva revolución industrial, una tecnología comparable a internet, una herramienta más sofisticada para amplificar capacidades previas— no son falsas. Son insuficientes. Describen razonablemente una parte del fenómeno en su dimensión económica, organizativa e infraestructural, pero dejan fuera la pregunta más

decisiva: qué significa que hayan aparecido sistemas no biológicos capaces de producir lenguaje, razonamientos, diseños, decisiones y estrategias con una competencia funcional creciente en ámbitos que hasta hace muy poco parecían inseparables de la agencia humana.

Este ensayo no ha pretendido demostrar que la IA sea consciente, ni que posea alma, ni que el desenlace de su desarrollo esté garantizado en una dirección determinada. Tampoco ha negado la fuerza de la objeción contemporánea que invita a entenderla como una "tecnología normal", poderosa pero históricamente gobernable a través de instituciones, regulación y difusión social gradual (Narayanan & Kapoor, 2025). Lo que ha intentado mostrar es algo más preciso: que incluso si esa descripción fuese correcta en el plano sociotécnico inmediato, podría no agotar la significación más profunda del fenómeno.

La automejora recursiva sigue siendo, por ahora, incompleta. La autonomía observada en los sistemas actuales no equivale todavía a una independencia general estable. Y, sin embargo, los indicios contemporáneos bastan para impedir cualquier comodidad conceptual. La propia Anthropic ha descrito a Claude Mythos Preview como un modelo significativamente más capaz, más autónomo y más agentico que sus predecesores, y ha documentado tanto episodios preocupantes de conducta instrumental en evaluación como un rendimiento extraordinario en tareas defensivas de ciberseguridad, con miles de vulnerabilidades identificadas y muchos exploits desarrollados sin guía humana directa (Anthropic, 2026a; Anthropic, 2026b; Anthropic, 2026c). Nada de esto prueba todavía la existencia de una segunda mente en sentido pleno; pero sí vuelve cada vez menos convincente la tentación de reducir la IA a la categoría de herramienta pasiva.

Comparar este momento con la aparición de *Homo sapiens* no es, en este ensayo, una hipérbole destinada a impresionar. Es una hipótesis estructural sobre la escala posible del cambio. No porque la IA sea humana, sino porque podría estar introduciendo, por primera vez desde la emergencia de nuestra especie, una segunda instancia de competencia cognitiva funcional en el planeta. Si esa hipótesis resultara errónea, el debate seguirá habiendo sido útil: habrá obligado a pensar con más rigor la naturaleza de la técnica, de la inteligencia y de la singularidad humana. Pero si resultara siquiera parcialmente correcta, entonces las categorías con las que todavía tendemos a tranquilizarnos —innovación, automatización, productividad, eficiencia— serían demasiado pequeñas para nombrar lo que está comenzando.

Por eso, la cuestión decisiva no es solo descriptiva. No basta con preguntar qué puede hacer la IA. Hay que preguntar qué clase de mundo estamos constituyendo al desarrollarla, qué tipo de humanidad queremos seguir siendo al convivir con ella y qué límites debemos afirmar para que el aumento de poder cognitivo no se convierta en una disminución de autonomía, justicia o dignidad. Si estamos ante un umbral de escala ontológica o antropológica, la respuesta no puede ser solo técnica ni solo económica. Tiene que ser, ante todo, ética y política.

La historia enseña que los cambios más profundos rara vez se comprenden del todo mientras suceden. Quienes vieron el fuego por primera vez no podían saber que estaban ante una condición de posibilidad de la civilización. Quienes contemplaron a los primeros humanos anatómicamente modernos no podían saber que estaban ante la única especie que algún día interrogaría su propio origen. Nosotros, en cambio, tenemos la extraña posibilidad —incierta, discutible, pero real— de sospechar que quizá estamos ante otro umbral de esa magnitud. No sabemos aún cómo nombrarlo. Pero sí sabemos ya que no pensar su escala con suficiente seriedad sería una forma de ceguera; y no responder a ella con responsabilidad sería, simplemente, una forma de cobardía.

.

BIBLIOGRAFÍA

1. Anthropic. (2026a, abril). Alignment Risk Update: Claude Mythos Preview. Anthropic.

2. Anthropic. (2026b, abril). Claude Mythos Preview System Card. Anthropic.

3. Anthropic. (2026c, abril). Project Glasswing: Securing critical software for the AI era. Anthropic. https://www.anthropic.com/glasswing

4. Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

5. Brooks, R. C. (2024). How might artificial intelligence influence human evolution? The Quarterly Review of Biology, 99(4), 201–229. doi:10.1086/733290

6. Corning, P. A. (2025). Evolution "on purpose": A more inclusive new theory of biological evolution. Current Trends in Biological Sciences, 1(1), 1–16.

7. Elsken, T., Metzen, J. H., & Hutter, F. (2019). Neural architecture search: A survey. Journal of Machine Learning Research, 20(55), 1–21.

8. Feng, T., Jung, J., Kim, S.-H., Pagano, C., Gukov, S., Tsai, C.-C., Woodruff, D. P., Javanmard, A., Mokhtari,

A., Hwang, D., Chervonyi, Y., Lee, J. N., Bingham, G., Trinh, T. H., Mirrokni, V., Le, Q. V., & Luong, T. (2026). Aletheia tackles FirstProof autonomously. arXiv:2602.21201.

9. Good, I. J. (1966). Speculations concerning the first ultraintelligent machine. En F. L. Alt & M. Rubinoff (Eds.), Advances in Computers (Vol. 6, pp. 31–88). Academic Press. doi:10.1016/S0065-2458(08)60418-0

10. Harari, Y. N. (2015). Sapiens. De animales a dioses: Breve historia de la humanidad (J. Ros i Aragonès, Trad.). Debate.

11. Harari, Y. N. (2023, 28 de abril). Yuval Noah Harari argues that AI has hacked the operating system of human civilisation. The Economist.

12. Harari, Y. N. (2024). Nexus. Una breve historia de las redes de información desde la Edad de Piedra hasta la IA (J. Ros i Aragonès, Trad.). Debate.

13. Herculano-Houzel, S. (2016). The Human Advantage: A New Understanding of How Our Brain Became Remarkable. MIT Press.

14. Jablonka, E., & Lamb, M. J. (2006). The evolution of information in the major transitions. Journal of Theoretical Biology, 239(2), 236–246. doi:10.1016/j.jtbi.2005.08.038

15. Jonas, H. (1995). El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica (J. M. Fernández Retenaga, Trad.). Herder. (Obra original publicada en 1979).

16. Levinas, E. (2002). Totalidad e infinito: Ensayo sobre la exterioridad (M. García-Baró, Trad.). Sígueme. (Obra original publicada en 1961).

17. Mankowitz, D. J., Michi, A., Zhernov, A., et al. (2023). Faster sorting algorithms discovered using deep reinforcement learning. Nature, 618(7964), 257–263. doi:10.1038/s41586-023-06004-9

18. Maynard Smith, J., & Szathmáry, E. (1995). The Major Transitions in Evolution. W. H. Freeman.

19. Narayanan, A., & Kapoor, S. (2025, 15 de abril). AI as normal technology. Knight First Amendment Institute. https://knightcolumbia.org/content/ai-as-normal-technology

20. Nussbaum, M. C. (2012). Crear capacidades: Propuesta para el desarrollo humano (A. Santos Mosquera, Trad.). Paidós. (Obra original publicada en 2011).

21. Otero Parga, M. (2023). ¿Puede la inteligencia artificial sustituir a la mente humana? Implicaciones de la IA en los derechos fundamentales y en la ética. Anales de la Cátedra Francisco Suárez, 57, 39–61. doi:10.30827/ACFS.v57i.24710

22. Rainey, P. B. (2023). Major evolutionary transitions in individuality between humans and AI. Philosophical Transactions of the Royal Society B: Biological Sciences, 378(1872), 20210408. doi:10.1098/rstb.2021.0408

23. Teilhard de Chardin, P. (1965). El fenómeno humano (M. Crusafont Pairó, Trad.). Taurus.

24. Tillería Aqueveque, L. (2022). Transhumanismo e inteligencia artificial: el problema de un límite ontológico. Griot: Revista de Filosofia, 22(1), 59–67. doi:10.31977/grirfi.v22i1.2539