

IA: una cartografía del nombre

Dr. Román García Fernández. Oviedo (España).

Alberto Fernández Fernández. Oviedo (España).

Recibido 28/06/2025 • Aceptado 28/10/2025

Resumen

La «inteligencia artificial» es ante todo un término confuso. El artículo recorre la utilización histórica de los términos *inteligencia* y *artificial* —de Platón y Aristóteles a Kant, Heidegger y Simondon—, argumentando que *artificial* funciona como concepto conjugado respecto de *natural*, y que la etiqueta «IA» sufre una inflación semántica al mezclar automatización, simulación y agencia. Siguiendo los planteamientos de Turing, se critica el desplazamiento hacia la cuestión ontológica, lo que favorece opacidades técnicas y concentraciones de poder bajo el rótulo unificador de «IA». Se confunden así niveles de análisis distintos, como muestra el caso «Sophia», ejemplo paradigmático de la ambigüedad entre persona jurídica y sujeto moral. Los autores proponen cartografiar el nombre, distinguiendo familias tecnológicas y sustituyendo «IA» por *Sistemas Asistidos por Ordenador (SAO)* y *Sistemas de Aprendizaje Autónomo / Control Delegado (MAA/CoDe)*, con el fin de clarificar responsabilidades, límites y riesgos de los protocolos de aprendizaje autónomo o control delegado.

Palabras clave: inteligencia artificial; nominalismo tecnológico; concepto conjugado; idea vs. categoría; prueba de Turing; persona jurídica vs. sujeto moral; control delegado (CoDe).

Abstract

«Artificial intelligence» is, above all, a confusing term. The article traces the historical use of *intelligence* and *artificial*—from Plato and Aristotle to Kant, Heidegger, and Simondon—arguing that *artificial* operates as a conjugate concept to *natural*, and that the label «AI» suffers from semantic inflation by conflating automation, simulation, and agency. Following Turing's approach, the paper criticizes the shift toward an ontological question, which enables technical opacity and concentrations of power under the unifying name of «AI». This conflation of distinct analytical levels is exemplified by the case of «Sophia», which illustrates the ambiguity between legal personhood and moral subjecthood. The authors propose mapping the name, distinguishing technological families and replacing «AI» with *Computer-Assisted Systems (CAS)* and *Autonomous Learning / Delegated Control Systems (ALS/CoDe)*, in order to clarify responsibilities, boundaries, and risks associated with autonomous learning or delegated control protocols.

Keywords: artificial intelligence; technological nominalism; conjugate concept; idea vs. category; Turing test; legal person vs. moral subject; delegated control (CoDe).

IA: una cartografía del nombre

Dr. Román García Fernández. Oviedo (España).

Alberto Fernández Fernández. Oviedo (España).

Recibido 28/06/2025 • Aceptado 28/10/2025

Introducción

La primera cuestión que deberíamos abordar es si la inteligencia artificial es un tema filosófico, o más concretamente, si puede ser tratado desde la filosofía. No se trata simplemente de que los filósofos escriban sobre el tema —lo cual, sin duda, ocurre—, sino de si dicha reflexión conserva un estatuto filosófico propio, más allá de contenidos sociológicos, psicológicos, antropológicos o científicos.

En principio, ya Terencio justifica de forma general el interés de la filosofía por todos los asuntos humanos: «*Homo sum, humani nihil a me alienum puto*» (*Hombre soy, nada de lo humano me es ajeno*). Pero, por si fuera poco, Platón, fundador de la filosofía académica, señala en el *Parménides* que no hay ideas menores, y que incluso los cabellos, las uñas o la suciedad merecen la atención del filósofo (*Parm.* 130c–e). Si nada de lo humano es ajeno, lo relevante no es tanto el objeto, sino la perspectiva desde la que se aborda dicho objeto. Mi colega Alberto Fernández ha desarrollado esta idea a partir de la denominada *Metáfora del semáforo*, donde el «momento amarillo» advierte sobre la confusión frecuente entre el análisis categorial (propio de las ciencias) y el análisis ideacional (propio de la filosofía). Este segundo momento del método, gnoseológico o positivo, es necesario para la filosofía, pero insuficiente.

Si bien la filosofía —al menos desde una tradición idealista o trascendental— se caracteriza por su atención a las ideas y no a las categorías, estas ideas no se producen *ex nihilo*, sino que emergen en el seno de las disciplinas empíricas: psicología, etología, biología, derecho, etc. En este sentido, tanto *inteligencia* como *artificial* son términos

que, desarrollados en múltiples disciplinas y técnicas, han sido posteriormente objeto de tratamiento filosófico en diversos momentos históricos.

Aunque la noción de inteligencia artificial fue formulada explícitamente por primera vez en el siglo XX, sus dos componentes —«inteligencia» y «artificial»— remiten, por separado, a conceptos filosóficos con una larga genealogía. La *inteligencia* ha sido objeto de reflexión desde Platón, para quien constituía la capacidad del alma racional de aprehender las Ideas, culminando en el conocimiento del Bien (*Rep.* 509d–511e). Aristóteles distinguirá entre un *nous poietikós* inmortal y activo, y un *nous pathetikós* pasivo y corruptible (*De Anima* III, 4–5; 429a10–430a25). Tomás de Aquino, en su *Summa Theologiae*, adapta esta distinción al pensamiento cristiano al proponer una separación entre *intellectus agens* e *intellectus possibilis* como facultades del alma racional (ST I, q.79, a.3–4). Descartes considera al ser humano una *res cogitans*, sustancia pensante separada del cuerpo (*Meditationes*, II), y en su *Traité de l'homme* plantea que el cuerpo, al margen del alma, puede entenderse como una máquina construida por Dios (AT XI, 120–130). Kant redefine la inteligencia como *entendimiento* (*Verstand*), la facultad de aplicar categorías puras a las intuiciones sensibles para generar conocimiento. En sus propias palabras: «Los pensamientos sin contenido son vacíos; las intuiciones sin conceptos, ciegas» (*KrV*, A51/B75). Más adelante sostiene que «el entendimiento es la facultad de las reglas» (A126), y que «la síntesis» constituye el fundamento de todo conocimiento (A77/B103).

En paralelo, lo *artificial* ha sido históricamente comprendido como producción técnica guiada por un principio externo. Aristóteles lo contrapone a la *physis* en su *Física* (II, 192b8–193b22). En la modernidad, Descartes considera el cuerpo como un autómatas (AT XI). Hobbes, en *Leviatán*, formula el Estado como un «hombre artificial», construido racionalmente para garantizar la paz (cap. XVII). Más recientemente, Heidegger, en *Die Frage nach der Technik*, advierte que la técnica moderna ya no es un simple instrumento, sino un modo de *desocultamiento* (*Entbergen*) en el que el ser es reducido a *fondo de reserva* (*Bestand*) mediante el *Gestell* (GA7, 16–25). Finalmente, Simondon propone una ontología del objeto técnico como ser en devenir, cuya génesis no debe reducirse a su utilidad, sino entenderse como un proceso de individuación (*Du mode d'existence des objets techniques*, Introducción y cap. 1).

Desde el punto de vista filosófico, conviene señalar una observación semántica: mientras que el término *inteligencia* se define frente a su negación —lo «no inteligente»—, el término *artificial* se conjuga en oposición a lo *natural*. Nos encontramos, así, ante lo que la lógica denomina un *concepto conjugado*, cuya relación con su opuesto puede adoptar diversas configuraciones lógicas: equivalencia ($A = N$), implicación ($A \subset B$), intersección parcial ($A \cap B \neq \emptyset$) o exclusión total ($A \cap B = \emptyset$). Desde nuestro punto de vista —y siguiendo una lectura histórico-dialéctica de la filosofía—, deberíamos partir de Kant como punto de inflexión teórica, dado que las posiciones anteriores han sido en gran parte superadas, lo que no impide que hayan dejado residuos conceptuales útiles para la reflexión actual.

¿De qué hablamos cuando hablamos de inteligencia artificial?

Conviene señalar que al hablar de *inteligencia artificial* no estamos simplemente ante una combinación conceptual como en el caso de *triángulo escaleno*. Esta última expresión remite a un concepto matemático bien delimitado: *triángulo* es un término definido dentro de una única categoría formal (la geometría euclidiana), y *escaleno* actúa como determinación interna, especificando un caso del género (*triángulo con lados desiguales*).

En cambio, *inteligencia artificial* no se compone de términos categoriales en sentido estricto, sino de lo que podríamos llamar ideas transversales o interdisciplinarias, que operan a través de múltiples marcos categoriales. «Inteligencia» es una noción que aparece en contextos muy diversos: psicología, neurociencia, filosofía de la mente, biología evolutiva, pedagogía, incluso en ámbitos militares. «Artificial», por su parte, no es una categoría positiva (como *ser vivo* o *triángulo*), sino una noción relacional, cuya inteligibilidad depende de su oposición a *natural*.

Desde esta perspectiva, *inteligencia artificial* no remite a un concepto unívoco, sino a una idea en el sentido kantiano, es decir, una representación que no puede ser dada empíricamente, pero que orienta la investigación o el pensamiento. No es una categoría científica cerrada, sino una constelación semántica en construcción, modelada por disciplinas diferentes que no siempre comparten los mismos marcos ontológicos o epistemológicos, ni tan siquiera metodológicos.

La diferencia es importante: mientras un concepto categorial como *triángulo escaleno* pertenece a un sistema cerrado de referencias definidas, una idea como *inteligencia artificial* permanece abierta, dinámica, y a menudo conflictiva, tanto en su definición como en su aplicación.

Diferencia entre concepto categorial e idea

Concepto categorial: un concepto es siempre categorial porque remite a una estructura formal delimitada por reglas internas de un campo disciplinar. Es cerrado, funcional y susceptible de definición operativa. Ejemplos:

- *Triángulo escaleno*: definido dentro de la geometría euclidiana.
- *Especie biológica*: en biología sistemática.
- *Mercado competitivo*: en economía.

Los conceptos categoriales operan dentro de sistemas de categorías científicas: *espacio, tiempo, cantidad, forma, función*, etc. Son susceptibles de falsación, cuantificación o modelización.

Idea: una idea es siempre transversal, entreverada, paradógica:

- No pertenece exclusivamente a una categoría científica.
- Se constituye en la intersección de múltiples disciplinas (psicología, neurociencia, informática, filosofía, derecho...).
- No puede ser definida unívocamente. Tiene validez orientadora, como las ideas kantianas de alma, mundo y Dios: guían la reflexión, sin agotarse en una experiencia empírica única (KrV, A327/B384).

Así, *inteligencia artificial* (IA) es más parecida a *progreso, vida, mente o cultura*, que a *electrón, fosfato, perihelio o ecosistema*.

Así, IA es más parecida a *progreso, vida, mente o cultura*, que a *electrón o ecosistema*.

«Artificial» como concepto conjugado

La observación de que el concepto «artificial» se define por oposición a «natural» es filosóficamente densa, problemática. Se trata de lo que algunos lógicos han llamado *conceptos conjugados*: su contenido depende del término contrario, sin formar una oposición simétrica clara. Vamos a sistematizarlo con la lógica de clases:

Relación	Ejemplo	Lectura
$A = N$	Todo lo artificial es natural (inclusión total)	Tecnologías biomiméticas, por ejemplo.
$A \subset N$	Lo artificial es un subconjunto de lo natural	Artificial como producción humana dentro del mundo natural.
$A \cap N \neq \emptyset$	Artificial y natural pueden coincidir parcialmente	Cultivos genéticamente modificados: creados, pero vivos.
$A \cap N = \emptyset$	Oposición excluyente	Tradición cartesiana: la máquina como no-vida, lo natural como lo que tiene alma.

Esto nos obliga a precisar qué entendemos por «natural» en cada caso: ¿lo no humano?, ¿lo espontáneo?, ¿lo biológico?, ¿lo que no ha sido diseñado? Sin esta aclaración, la distinción *artificial/natural* puede volverse ambigua, incluso ideológica. En este sentido, José Ignacio Galparsoro en *Mas allá del posthumanismo. Antropotécnicas en la era digital* (2019), destruye el concepto de «mejoramiento humano» propuesto por el posthumanismo, apegado exclusivamente al avance biomédico o cibernético, al considerar que la lectura o la escritura son procesos de mejoramiento y modificación humana con siglos de historia, a los que cabría sumar muchas otras prácticas humanas.

Inteligencia artificial y otros conceptos transversales

Para mostrar de forma fehaciente que estamos ante una idea plural, puede ser útil comparar la IA con otros conceptos similares:

Idea transversal	Campos implicados	Carácter filosófico
Vida	Biología, medicina, filosofía, derecho	¿Qué es estar vivo? (desde Aristóteles hasta Maturana)
Conciencia	Psicología, neurociencia, fenomenología	¿Qué es tener experiencias subjetivas?
Máquina	Ingeniería, ontología, cibernética	¿Qué es lo que hace que algo sea una máquina? (Deleuze y Guattari, Simondon)
Lenguaje	Lingüística, filosofía analítica, IA	¿El lenguaje es exclusivo del humano?, ¿puede una máquina «entender»?
Inteligencia	Psicología, IA, educación, filosofía de la mente	¿Es inteligencia sin conciencia realmente inteligencia?

La IA recoge todas estas cuestiones, y por tanto funciona como un *concepto frontera* o *nodo semántico*, más que como una definición única. Esta condición es lo que le otorga interés filosófico.

¿Por qué «una cuestión de nombre»?

Plantear que el nombre *inteligencia artificial* opera como un significante totalizante, vago y polisémico es una crítica potente que toca el núcleo semántico y epistemológico del debate contemporáneo.

La expresión *inteligencia artificial* se aplica hoy a una gran diversidad de fenómenos: desde un asistente de voz hasta un sistema de predicción militar o una red neuronal generativa. Al nombrar todo bajo una misma etiqueta, se borra la diferencia entre lo mecánico, lo cognitivo, lo emocional y lo social.

Si el término IA sirve para ocultar o legitimar procesos muy distintos — automatización laboral, control policial, marketing personalizado, simulación afectiva —, entonces se convierte en un concepto ideológico: nombra tan solo para fascinar, para epatar o, peor aún, para desinformar, para mentir.

Con el modismo IA se confunden *ideas* y *categorías*, y lo que es más grave, existe una confusión entre *lo que se piensa* y *lo que se nombra*, lo que impide claramente llegar a un punto de entendimiento. Nos encontramos ante una situación mistificadora en la que el término AI se convierte en el *mito totalizante* de la *inteligencia artificial*. Convendría

llamar la atención sobre lo que decía Platón en el *Cratilo*: nombrar es ordenar el mundo, pero también es ocultar la diferencia bajo una unidad ficticia.

Por todo ello nos encontramos ante una situación en la que el término AI se convierte en el mito totalizante de la «inteligencia artificial».

Por una parte, quiero llamar la atención sobre lo que decía Platón en el *Cratilo*: nombrar es ordenar el mundo, pero también es ocultar la diferencia bajo una unidad ficticia.

¿Puede pensar una máquina?

En su célebre artículo de 1950 *Computing Machinery and Intelligence*, Alan Turing no preguntaba literalmente si una máquina puede pensar. De hecho, afirmaba que esa pregunta estaba mal formulada y difícilmente podría tratarse en términos rigurosos. Según Turing, el problema no es metafísico ni psicológico, sino funcional: lo relevante no es si una máquina *piensa* en sentido humano, sino si puede comportarse de tal modo que un observador no pueda distinguirla de un ser humano en una conversación escrita.

Propongo considerar la pregunta: ‘¿Pueden pensar las máquinas?’ Como primer paso, habría que definir el significado de los términos ‘máquina’ y ‘pensar’. Pero, en lugar de intentar tal definición, reemplazaré la pregunta por otra, más concreta y menos ambigua. Turing, *Computing Machinery and Intelligence* (1950)

La operación que propone Turing —hoy conocida como el *test de Turing*— consiste en sustituir una pregunta ontológica por un experimento de simulación conversacional. Este enfoque ha sido muchas veces malinterpretado o desvirtuado. No se trata de afirmar que la máquina *piense* como un humano, sino de preguntar si puede generar una apariencia funcional de pensamiento humano en un contexto comunicativo.

La película *Ex Machina* (Alex Garland, 2015) ofrece una dramatización contemporánea particularmente fiel a este espíritu. El protagonista no debe verificar si la androide Ava *piensa, siente o comprende*, sino si logra convencerle —en una

interacción libre y no guiada— de que no se trata simplemente de un simulacro. El test no consiste en comparar su inteligencia con la humana, sino en evaluar la capacidad de la simulación para producir indistinción.

Sin embargo, muchas lecturas de este experimento, tanto populares como académicas, ignoran este giro metodológico de Turing, y proyectan en su propuesta una pregunta que él mismo había considerado filosóficamente estéril. Esta confusión inicial ha contribuido a cargar el término *inteligencia artificial* de una ambigüedad semántica persistente. La distinción, ya clásica, entre *IA débil* (limitada a tareas específicas) e *IA fuerte* (general, autoconsciente o intencional), no hace sino replicar el malentendido fundacional. Lo mismo ocurre con expresiones como *IA actual* frente a *IA futurible*, o *inteligencia artificial generativa* frente a *conciencia artificial*: el uso del mismo nombre para realidades radicalmente distintas oscurece más que clarifica.

La inflación semántica de la IA

El uso del término *inteligencia artificial* en el discurso contemporáneo presenta un fenómeno de inflación semántica: se aplica a objetos, procesos y sistemas radicalmente distintos entre sí. Desde un asistente doméstico que reconoce órdenes de voz, un coche autónomo, hasta una red neuronal capaz de generar imágenes complejas, pasando por algoritmos de diagnóstico médico o sistemas de recomendación comercial, todo se etiqueta como «IA». Se llama «inteligente» tanto al cálculo optimizado como al lenguaje predictivo, tanto al automóvil robótico como al *chatbot* afectivo, como si la mera rúbrica «IA» aportara una cualidad sustancial.

Esta generalización ha dado lugar a una nube conceptual confusa, en la que se mezclan desarrollos reales, expectativas futuristas, simulaciones superficiales y modelos aún inexistentes. Como resultado, se crea una imagen unificada y espectacular de algo que, en realidad, es profundamente heterogéneo, se borran distinciones importantes y se genera un efecto de fetichización tecnológica. De esta forma, la IA se convierte en una categoría ideológica que unas veces se asocia a progreso, inteligencia, humanidad futura... otras a una visión pesimista del mundo, distopías y paraísos perdidos, cuando en realidad sólo designa algoritmos estadísticos, minería de datos y lógica binaria.

Este fenómeno tiene un efecto filosófico evidente: al nombrar todo como IA, no se entiende nada con claridad. El término funciona como una metáfora desplazada, un significante totalizante que, en lugar de esclarecer, confunde. Nombrar todo como IA impide pensar con precisión: ¿hay conciencia? ¿hay intención? ¿quién decide? ¿qué tipo de inteligencia? ¿Quién tiene la responsabilidad? Se incurre en homonimia, lo mismo dicho de muchas cosas, pero sin identidad real, como señaló Aristóteles (*Metafísica*, IV, 2). Es, por tanto, necesario distinguir entre IA simbólica, estadística, generativa, cibernética, etc.

Por todo ello, necesitamos precisar una semántica del término, hacer una cartografía del nombre. Necesitamos distinguir: ¿qué tipos de IA hay? ¿qué significan en cada contexto? ¿cómo impactan? Nombrar bien para pensar mejor.

La ambigüedad interesada

La polisemia del término «inteligencia artificial» no es solo confusión epistemológica, lingüística: es también una estrategia discursiva y cumple una función política. Bajo este nombre se ocultan tecnologías muy distintas, intereses industriales, marcos regulatorios laxos y, sobre todo, una batalla por el control del desarrollo técnico. En muchos casos, la polisemia del término IA beneficia a quienes detentan el poder de nombrar y definir qué es y qué no es «inteligencia artificial». Grandes corporaciones tecnológicas —OpenAI, Google, Meta, Microsoft— invocan la necesidad de regular la IA en términos globales, pero con ello enmascaran que los sistemas que diseñan son productos centrípetos, cerrados, dirigidos, entrenados y sesgados.

Un caso paradigmático lo ofrece el proyecto DeepSeek, una IA de «código abierto» desarrollada con un presupuesto de tan solo 6 millones de dólares, que logró resultados comparables a los de modelos privativos de empresas como OpenAI o Google DeepMind. Este caso revela dos cosas: que la «necesidad» de inversiones estatales multimillonarias en IA puede ser una narrativa corporativa interesada; y que el desarrollo no es neutral, ya que el comportamiento de un sistema de IA depende de los datos que se le den, las instrucciones que reciba, y los objetivos que se le asignen.

En un modelo abierto (DeepSeek), esto puede auditarse; en uno cerrado (OpenIA), la transparencia es absolutamente inviable.

Los sistemas de IA de «código cerrado» no permiten examinar los datos ni las reglas de entrenamiento, lo que limita la transparencia y la capacidad de crítica externa. Así, la denominación «inteligencia artificial» funciona también como mecanismo de opacidad. Bajo su prestigio discursivo se ocultan procesos algorítmicos de minería de datos, estructuras jerárquicas de entrenamiento estadístico y decisiones de diseño cargadas de implicaciones culturales, ideológicas y económicas. Nombrar todo esto como si fuera una «inteligencia» homogénea y objetiva es una forma de naturalizar lo que en realidad es un artefacto dirigido. La IA no es solo una técnica, es una decisión política y económica, es una imposición epistémica.

Sophia es una persona... pero no humana

En otro artículo de esta misma Revista desarrollamos la perspectiva de la AI desde el punto de vista ético y desarrollamos dos cuestiones muy interesantes, sobre todo a partir del proyecto Sophia, un robot humanoide al que se le ha concedido la nacionalidad saudí. La primera cuestión que se nos viene a la cabeza es que, al igual que Pinocho, Sophia no tiene sexo; y la segunda que puede mentir, como Pinocho.

Sophia, el *ginoide* de aspecto femenino desarrollado por Hanson Robotics, fue reconocido oficialmente en 2007 como ciudadana por Arabia Saudí. Desde entonces, múltiples artículos y debates se han referido a ella como «persona», a menudo sin especificar en qué sentido.

Empero, afirmar que Sophia «es una persona», lo que no implica necesariamente que sea una persona humana. Desde el punto de vista filosófico y jurídico, el concepto de «persona» admite distintas modulaciones que conviene distinguir cuidadosamente

Tipo de persona	Ejemplo	Criterio de atribución
Persona humana (natural)	Ser humano	Autonomía biológica, conciencia, voluntad
Persona jurídica (legal)	Empresa, Estado, fundación	Capacidad de actuar legalmente, representación
Persona simbólica o teológica	Padre, Hijo, Espíritu Santo	Relación trinitaria, sin corporeidad
Persona cibernética reconocida	Sophia (Arabia Saudí, 2017)	Acto político-mediático, sin derechos plenos

Desde esta perspectiva, Sophia no es en modo alguno una *persona humana*, ni posee las condiciones materiales o fenomenológicas que la filosofía ha asociado tradicionalmente con la subjetividad: no tiene cuerpo vivido, ni conciencia de sí, ni historia personal, ni puede establecer relaciones éticas auténticas. Pero eso no impide que un estado soberano le haya atribuido el estatus de «persona» en sentido jurídico o simbólico, de forma semejante a como lo hace con empresas o entidades colectivas.

Este caso pone de manifiesto que el uso público del término «persona» no siempre refleja una realidad ontológica, sino que cumple funciones políticas, estratégicas o discursivas. En el caso de Sophia, el reconocimiento de su personalidad parece tener más valor comunicativo y especulativo que práctico: no está claro qué derechos u obligaciones reales derivan de dicha ciudadanía artificial.

En un contexto dominado por la retórica de la *inteligencia artificial*, el uso indiscriminado de términos como *persona*, *conciencia* o *autonomía* corre el riesgo de naturalizar categorías que deberían ser sometidas a examen crítico. Sophia es una *persona jurídica* del mismo modo que lo son una universidad o una compañía minera, pero eso no la convierte en sujeto moral ni en inteligencia autónoma.

De la categoría a la responsabilidad

El caso de Sophia ejemplifica con claridad lo que hemos venido defendiendo: el uso confuso y expansivo de ciertos términos como «inteligencia», «persona» o «autonomía» no es meramente un problema lingüístico, sino filosófico y político. Cuando se aplica el mismo nombre a fenómenos radicalmente distintos —una conciencia viviente, una empresa multinacional, una entidad divina o un sistema de

conversación robótica—, lo que se produce no es una extensión del conocimiento, sino una suspensión de la crítica.

Atribuir personalidad jurídica o nacionalidad a una máquina no equivale a reconocerle conciencia ni dignidad, pero sí introduce elementos de ambigüedad en el debate público, jurídico y ético. Por ello, se hace imprescindible distinguir entre *nombres funcionales* y *nombres fundados*, entre el estatus que se concede a un ente o entidad y la sustancia que dicho ente o entidad posee.

En el artículo referido con anterioridad, “Geppetto moldea a Pinocchio”, que completa este escrito, desarrollamos con más profundidad las implicaciones éticas de esta operación simbólica, a través de dos interrogantes especialmente reveladores inspirados en el mito de Pinocho:

- ¿Puede ser considerado sujeto quien no tiene sexo ni historia?
- ¿Es lícito diseñar máquinas que simulen sentimientos o emociones, o más aún, que mientan estratégicamente?

76

Estas preguntas, lejos de ser anecdóticas, nos devuelven al núcleo filosófico de la técnica: el de su capacidad para configurar lo humano a través de la apariencia.

La IA desde el espacio antropológico

Una de las complejidades de la utilización del término AI es las distintas dimensiones sobre las que se despliega, por ello cabría aplicar la teoría del espacio antropológico (Bueno, 1978), lo que nos permitiría delimitar en que campo y a qué nivel de relaciones estamos aludiendo. Esbozando brevemente su aplicación, podríamos señalar que desde el eje radial, nos encontramos con artefactos, herramientas, máquinas y materiales que no plantean otros problemas que los de otras herramientas. Los cambios tecnológicos producen procesos de innovación y modificaciones como las que supusieron el arado de reja o la máquina de vapor. Mientras que en el eje circular, la IA entra en el plano de las relaciones humanas, los derechos, tanto de las máquinas como referidas a los propios hombres (derecho a la intimidad, *face news*, ...). Por último, para no extenderme más, el eje angular nos sitúa

ante el problema de la conciencia. Nos sitúa ante algo, que como los animales, los dioses o los extraterrestres, funcionan como un algo que no nosotros y que reacciona o interactúa con nosotros. En definitiva, este problema es equiparable al de la inteligencia animal o a una supuesta inteligencia divina o extraterrestre.

Nombrar para pensar

En este contexto, recuperar una crítica filosófica del nombre se vuelve urgente. No se trata de buscar una definición esencial de la inteligencia artificial —como si esta pudiera ser especificada de forma aislada—, sino de descomponer el término, clarificar sus usos y cartografiar sus funciones.

Esto implica distinguir, al menos:

- Entre IA funcional y IA simbólica.
- Entre simulación conductual y comprensión semántica.
- Entre herramienta programada y agente que aprende autónomamente.
- Entre sistema cerrado y sistema abierto y auditado.

Sólo así será posible evitar tanto la fetichización como el catastrofismo, y pensar la inteligencia artificial no como un mito totalizador, sino como un campo técnico, epistémico y político que debe ser analizado con precisión.

Comparativa: Tecnologías frecuentemente confundidas con «inteligencia artificial»

Tecnología / Ejemplo	Categoría tecnológica	¿Es IA propiamente dicha?	Notas aclaratorias
Robótica (Asimov)	Robótica / Simulación	No	Puede incorporar IA, pero no lo es por sí misma
Telares Jacquard (1801)	Mecanización	No	Sistema mecánico, sin computación
Ford T – cadena de producción (1908)	Mecanización humana	No	Involucra trabajo humano en estructura repetitiva
Automatización Toyota	Automatización avanzada	Parcialmente	Automatiza sin necesidad de IA
Prótesis	Capacidad mecánica aumentada	No	Extiende capacidades físicas humanas
Exoesqueletos	Capacidad mecánica aumentada	No	Asistencia mecánica sin toma de decisiones
Biotecnología	Ingeniería biológica	No	Manipulación genética, no computacional
Realidad aumentada (reconocimiento facial, matrículas)	Interfaz / Reconocimiento	No	Utiliza algoritmos, pero no implica comprensión
Realidad virtual	Simulación / Entorno virtual	No	No implica procesamiento autónomo
Realidad inmersiva	Entorno inmersivo	No	Puede simular IA, sin ser IA
Transhumanismo (Nick Bostrom)	Filosofía tecnológica / Prospectiva	No	Tesis filosófica, no tecnología concreta
Singularidad tecnológica	Filosofía tecnológica / Futurismo	No	Hipótesis sobre IA futura, no actual
Machine learning	IA débil / Aprendizaje automático	Sí (IA estadística)	Aprende patrones sin comprender
Big Data	Gestión de datos masivos	No	Soporte para IA, no IA en sí misma
Publicidad basada en datos	Gestión de datos / Targeting	No	No hay deliberación ni inteligencia
Modificación de conductas (ej. elecciones)	Psicología digital / Neurocognición aplicada	Parcialmente	IA usada para influir conductualmente
Sistema experto (análisis de datos)	IA simbólica	Sí	Procesa datos según reglas definidas
Sistema experto (toma de decisiones)	IA simbólica	Sí	Ejecuta decisiones dentro de un marco lógico
Sistemas CoDe (control delegado) (Marcos, 2023)	Gobernanza algorítmica / Crítica de la IA	Depende del diseño	Proponemos sustituir «IA» por «control delegado»

El planteamiento de Carlos Madrid sobre la IA en su obra *Filosofía de la inteligencia artificial* como mero tratamiento masivo de datos se articula en una crítica a lo que él considera el mito de la inteligencia y el mito de lo artificial. Esta visión puede ser muy útil para analizar las categorías propuestas en la tabla de clasificación tecnológica precedente.

Madrid sostiene que lo que hoy se llama «inteligencia artificial» no es ni inteligencia ni artificial, en el sentido filosófico fuerte. Según él, se trata más bien de un conjunto de tecnologías estadísticas y algorítmicas orientadas al procesamiento de datos, y no de una instancia que posea *inteligencia* en el sentido humano, abductivo y situado.

Tabla, agrupando las tecnologías según el criterio de Madrid Casado y destacando los puntos de fricción:

Categoría	Tecnología	¿IA según Madrid?	Comentario crítico (Madrid)
Automatización	Telares de Jacquard, Ford T, Toyota	✗ No	No hay «inteligencia», solo ejecución de tareas repetitivas.
Capacidades mecánicas aumentadas	Prótesis, exoesqueletos	✗ No	Tecnología asistencial, sin decisiones ni procesamiento de sentido.
Robótica	Robots industriales, autónomos	⚠ Parcialmente	Inteligencia sintáctica y predictiva, pero sin comprensión semántica ni corporalidad.
Biotechnología	Ingeniería genética, implantes	✗ No	Es técnica médica, no una inteligencia autónoma.
Realidad aumentada/virtual	Reconocimiento facial, realidad inmersiva	⚠ Parcialmente	Procesamiento masivo de imágenes; no hay inferencia ni propósito.
Machine Learning / Big Data	Modelos predictivos, clasificación	✓ Sí (como IA real existente)	Pero reducida a inducción estadística, sin abducción ni sentido.
Gestión de datos / Publicidad	Sistemas de segmentación, elecciones	✓ Sí (pero delegada)	El sistema no «decide», sino que aplica patrones definidos por humanos.
Sistemas expertos	Diagnóstico médico, predicción legal	⚠ Parcialmente	Son reglas codificadas; inteligencia exógena (no autónoma).
Control delegado (CoDe)	Marcos (2023)	✓ Como redefinición honesta	Se propone abandonar el término IA y hablar de sistemas de <i>control delegado</i> .
Singularidad / Transhumanismo	Nick Bostrom, cíborgs, mente digital	✗ No (según Madrid)	Ideas escatológicas, basadas en mitos (IA fuerte, transferencia de mente).

Sin embargo, frente a la visión de Madrid, hoy existen desarrollos que van en otra dirección, y los podemos agrupar como sistemas ciberfísicos con bucles de

percepción–acción–aprendizaje–decisión. Hoy existen máquinas que se caracterizan por:

1. Recepción de estímulos: sensores, visión artificial, audio, proximidad, etc.
2. Aprendizaje adaptativo: no solo entrenados previamente, capaces de reajustarse en tiempo real.
3. Toma de decisiones: no seguir reglas fijas, elegir entre alternativas (con margen de indeterminación).
4. Interacción en el mundo: retroalimentación (*feedback*) continua del entorno.
5. Control delegado, en el sentido que propone Broncano o Marcos (2023): sistemas a los que se transfiere parte del control operativo o decisional.

Algunos ejemplos de tecnologías que se aproximan a esto

Robots autónomos con aprendizaje reforzado como: robots de Boston Dynamics + entornos de aprendizaje como OpenAI Gym + Deep Reinforcement Learning.

- Reciben datos sensoriales.
- Aprenden estrategias motoras por prueba y error (no solamente por correlación estadística).
- Pueden adaptarse a entornos no conocidos previamente.

No solo procesan datos sino que exploran, evalúan y actúan sobre el entorno, aunque aún sin comprensión semántica o motivación propia.

Vehículos autónomos (nivel 4-5) como: Waymo, Tesla (en teoría), proyectos de Nvidia, etc.

- Integran visión artificial, predicción de trayectorias, planificación, aprendizaje en tiempo real.
- Toman decisiones de navegación y reacción a eventos externos.

Aunque críticas como las de Carlos Madrid dirían que todo sigue siendo procesamiento sintáctico y optimización de rutas, hay margen real de indeterminación

e inteligencia en situaciones complejas, lo que se acerca a lo que se denomina «control delegado».

Agentes en entornos de simulación con exploración activa como AlphaGo Zero o AlphaStar (DeepMind)

- Aprenden *desde cero*, sin datos humanos.
- Generan su propio conocimiento estratégico a través de millones de interacciones en entornos simulados.
- Eligen sus movimientos y tácticas sin intervención humana directa.

En definitiva, no están relacionados con el mundo físico, lo que sería un hándicap importante, pero son sistemas con una forma real de autonomía decisional.

Más interesante es la IA embebida en cuerpos: robótica cognitiva o enológica como por ejemplo iCub o NICO, o robótica inspirada en Varela y Brooks (*robotics without representation*).

- No se basan en una representación interna del mundo, sino en acoplamientos dinámicos.
- Aprenden a actuar mediante interacción, no desde modelos preestablecidos.

Aquí hay un giro epistémico: no se trata de procesar datos, sino de participar en una dinámica de mundo, aunque de forma muy rudimentaria aún.

Estos desarrollos, aunque aún distantes de una «inteligencia fuerte», desafían la visión puramente estadística de la IA. Plantean un escenario intermedio: sistemas sin conciencia ni comprensión, pero con inteligencia funcional y autonomía relativa. Por tanto, más que negar toda forma de inteligencia, tal vez convenga repensar lo que entendemos por «inteligente» cuando se aplica a lo técnico.

Esquema comparativo de modelos de agencia artificial

Modelo tecnológico	Ejemplo representativo	Aprendizaje autónomo	Estímulos del entorno	Toma de decisiones no programadas	Autonomía funcional	¿IA fuerte ?
Procesamiento estadístico masivo	Publicidad programática, segmentación electoral	No	No	No	No	No
Sistema experto simbólico	Diagnóstico médico experto, software jurídico	No	No	No	Baja	No
Machine learning supervisado	Clasificadores de imágenes, chatbots tradicionales	Limitado	Sí	Limitada	Media	No
Aprendizaje reforzado (RL)	Agentes de juego, simuladores de movimiento	Sí	Sí	Sí	Alta	No
Agente en simulación (AlphaZero)	AlphaZero, AlphaStar (DeepMind)	Sí	Simulado	Sí	Alta	No
Vehículo autónomo (nivel 4-5)	Waymo, Tesla Autopilot (avanzado)	Sí	Sí	Sí	Alta	No
Robótica cognitiva	Robots humanoides (iCub, NICO)	Sí	Sí	Sí	Alta	No
Drones militares autónomos	Loitering munitions, Iron Dome	Sí	Sí	Sí	Alta	No
Sistemas CoDe (control delegado)	Propuesta conceptual (Marcos, 2023)	Variable	Depende del diseño	Sí	Variable	No

No hay que buscar una definición esencial de IA, sino clarificar sus usos y funciones. Cabría distinguir entre:

- IA funcional / simbólica
- Simulación / comprensión
- Herramienta / agente
- Código abierto / sistema cerrado

Solo así evitaremos caer en el mito totalizador de la IA.

Propuesta de Análisis

Tipo de máquina (con ejemplo)	Características principales	Movilidad	Estímulos del entorno	Procesamiento	¿Modifica su programa?	Nivel de autonomía
Ordenador clásico (Ej: PC de sobremesa)	Ejecuta instrucciones fijas, sin aprendizaje	No	Solo entrada humana (teclado, ratón)	Algoritmos deterministas	No	Nulo
Programa simbólico (Ej: ELIZA)	Simula diálogo sin comprensión	No	Entrada textual humana	Estructuras lingüísticas fijas	No	Nulo
Sistema experto (Ej: MYCIN)	Aplica reglas sobre bases de datos	No	Entrada estructurada (datos simbólicos)	Inferencia lógica basada en reglas	No	Bajo
Modelo generativo LLM (Ej: ChatGPT)	Genera texto con estadísticas y patrones	No	Bases de datos + input humano	Predicción estadística (<i>transformers</i>)	No (solo reentrenamiento externo)	Medio
Máquina funcional (Ej: impresora 3D, Roomba básica)	Cumple funciones físicas automatizadas	Parcial	«Mínimos (temperatura, comandos simples)»	Control programado	No	Bajo
Robot móvil sin IA (Ej: primer robot aspirador)	Se desplaza siguiendo rutas preprogramadas	Sí	«Rutas definidas, sin sensores reales»	Rutas y bucles fijos	No	Bajo
Robot con sensores y IA limitada (Ej: Spot de Boston Dynamics)	«Interacción con sensores, responde al entorno»	Sí	«Sensores básicos (cámara, lidar)»	Retroalimentación y ajuste reactivo	Parcialmente (ajustes internos)	Medio
Robot cognitivo/aprendizaje reforzado (Ej: iCub, Atlas)	«Aprende con experiencia, ajusta estrategias»	Sí	«Sensorial compleja (visual, háptica, espacial)»	Aprendizaje reforzado y continuo	Sí	Alto
Dron militar autónomo (Ej: <i>loitering munition</i> tipo Harpy)	Identifica y actúa sobre objetivos sin guía humana	Sí	«Sensores múltiples (imagen, térmico, localización)»	Evaluación táctica autónoma	Sí	Alto

Propuesta de términos para distinguir tecnologías

Para finalizar, proponemos evitar el nombre de Inteligencia Artificial y utilizar el más correcto de Sistemas asistido por ordenador y Sistemas de aprendizaje autónomo para referirnos a dos realidades bien diferenciadas.

Uso	Término	Justificación
Sistemas asistidos por ordenador, sean físicos o lógicos	SAO (<i>Sistemas Asistidos por Ordenador</i>)	Término técnico claro, amplio y no mitificante.
Máquinas que toman decisiones y modifican su programación	Sistemas de Aprendizaje Autónomo o MAA (<i>Máquinas con Inteligencia Artificial</i>)	Evita decir «inteligentes» o «conscientes». Pone énfasis en la capacidad de modificar comportamientos según experiencias, sin suponer subjetividad.

Bibliografía

- Aristóteles. (1994). *Metafísica* (T. Calvo Martínez, Trad.). Madrid: Gredos.
- Aristóteles. (2020). *Acerca del alma (De Anima)* (T. Calvo Martínez, Trad.). Madrid: Gredos.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.
- Bueno, G. (1975 y ss.). *Teoría del cierre categorial*. Oviedo: Pentalfa Ediciones.
- Bueno, G. (1978). «Sobre el concepto de ‘espacio antropológico’», *El Basilisco* 5, 57-69.
- Bueno, G. (1996). *El sentido de la vida*. Oviedo: Pentalfa Ediciones.
- Galparsoro, J. I. (2019). *Más allá del posthumanismo: Antropotécnicas en la era digital*. Granada: Comares.
- Heidegger, M. (1954). *La pregunta por la técnica* (E. Barjau, Trad.). Barcelona: Ediciones del Serbal.
- Jonas, H. (1979). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica*. Barcelona: Herder.
- Kant, I. (2009). *Crítica de la razón pura* (P. Ribas, Trad.). Madrid: Gredos.
- López Brugos, J. A. (1989). «Los chimpancés mecánicos». *El Basilisco*, 7, 65-78. Oviedo: Fundación Gustavo Bueno.
- Madrid Casado, C. M. (2024). *Filosofía de la inteligencia artificial*. Oviedo: Pentalfa Ediciones.
- Marcos, J. (2023). «Propuesta de sistemas de control delegado». *Revista de Filosofía y Tecnología*. [Datos editoriales no verificados].
- Platón. (1997). *La República* (C. Eggers Lan, Trad.). Madrid: Gredos.
- Platón. (1998). *Parménides* (J. A. Míguez, Trad.). Madrid: Gredos.
- Searle, J. R. (1980). «Minds, brains, and programs». *Behavioral and Brain Sciences*, 3(3), 417-424. <https://doi.org/10.1017/S0140525X00005756>
- Simondon, G. (1958). *Du mode d’existence des objets techniques*. Paris: Aubier.
- Turing, A. M. (1950). «Computing machinery and intelligence». *Mind*, 59(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>

Otros recursos

Ex Machina, dir. Alex Garland (2015)

DeepSeek (2024), proyecto de IA abierta

Hanson Robotics, proyecto Sophia

Robots inteligentes Toyota

<https://www.toyota.es/world-of-toyota/articles-news-events/robots-toyota-movilidad>

