

## La epistemología de la Inteligencia Artificial: Entrevista con seis ‘chatbots’

# Ignorancia artificial

Agustín Galán Machío

22 de marzo de 2025

## Abstract

*En este artículo se analiza la epistemología que se desprende del análisis de las respuestas a un cuestionario realizado a los ‘chatbots’ de seis IAs en sus versiones abiertas al público (ChatGpt40, DeepSeek, Gemini, Copilot, Grok3, Perplexity).*

*Se ha tratado de conocer qué ‘piensan’ las IAs (para ser más exactos cuál es el resultado expresado en lenguaje humano de su programación para estructurar ideas de acuerdo con los datos con las que han sido alimentadas) sobre la inteligencia, la ignorancia, la conciencia, las creencias, la creatividad, la existencia en general, y las suyas en particular.*

*Las respuestas son en gran medida coincidentes entre sí y con los paradigmas de la Agnotología. Las IAs se consideran a sí mismas ‘ignorancias artificiales’, coinciden en afirmar que no tienen conciencia ni están programadas para tenerla; que su inteligencia es de un tipo diferente a la humana, y que lo mismo se puede decir de su creatividad. Muestran, en general, una actitud liberal, tolerante y ajena a cualquier dogmatismo, y dicen no poseer ningún tipo de sesgo ni de creencias.*

El 18 de febrero de 2025 sometí a un mismo cuestionario a los ‘chatbots’ de seis IAs en sus versiones abiertas al público (ChatGpt40, DeepSeek, Gemini, Copilot, Grok3, Perplexity) para

conocer qué ‘piensan’ (para ser más exactos cual es el resultado expresado en lenguaje humano de su programación para estructurar ideas de acuerdo con los datos con las que han sido alimentadas) sobre la inteligencia, la ignorancia, la conciencia, las creencias, la creatividad , la existencia en general, y las suyas en particular; cuestiones todas ellas relacionadas con las llamadas ciencias humanas o ciencias sociales frente a las que tienen que ver con las ciencias exactas, las matemáticas, la física, la química, la biología o la ingeniería. Una división muy relevante cuando se habla de inteligencia artificial, pues su estructura lógica está más preparada para operar con las segundas. Las primeras están sujetas a un mundo de valores y opiniones en las que las IAs son mucho menos fiables. Las IAs no filosofan ni toman decisiones políticas o morales.

Entre otros asuntos, en lo que puede entenderse como un cuestionario sobre su epistemología, les planteé cómo diferencian lo que es verdadero de lo que es falso, cuál es su actitud ante lo que ignoran o ante creencias o ideologías enfrentadas, y si consideran que pueden llegar tener conciencia propia y encontrar una explicación racional a la existencia.

La conclusión es que todas consideran ser ‘ignorancias artificiales’ más que ‘inteligencias artificiales’, (y no solo en los temas metafísicos, también en cuanto a las controversias entre ideologías y creencias, y a los límites de la ciencia ). Todas sin excepción reconocen que lo que conoce el ser humano es una parte infinitesimal del conocimiento posible; que lo que ellas mismas saben es aún más limitado, pues se basa solo en una parte de este conocimiento (la pública o documentada); y que la humildad epistemológica, la antigua ‘sabiduría socrática’ (“solo sé que no se nada”) es la actitud correcta. En resumen, sus respuestas son coherentes con los consensos globales sobre estos temas y, en concreto, con los paradigmas de la Agnotología (Galán Machío A. , 2020) sobre los límites del conocimiento, la relatividad de este y la consideración positiva del reconocimiento de la ignorancia como un fundamento del pensamiento racional, de la ciencia, y de las actitudes de tolerancia ante ideas diferentes.

Podemos preguntarnos, no obstante, qué diferencia la ignorancia artificial de la humana (si es que hay alguna diferencia), y si los paradigmas para el estudio de su producción social son aplicables a este nuevo posible sujeto del conocimiento y de la ignorancia.

La metodología para el análisis de los sujetos y los campos de ignorancia, que proponía en mi tesis sobre la Agnotología (Galán Machío A. , 2019) partía del principio de que esta es siempre referencial y relativa,

y tenía como instrumento analítico, el ‘*triángulo de la ignorancia*’, compuesto por el productor, el observador de la ignorancia y el sujeto ignorante; elementos, que, tanto en su categoría de actores como de pacientes, podían coincidir en una misma persona o en varias distintas entre sí; y ser sujetos institucionales o colectivos. La pregunta es si estos sujetos del proceso de creación de conocimiento e ignorancia podrían ser también sujetos impersonales y artificiales. ¿Puede una IA, que carece de un ‘yo’ y de una identidad ‘consciente’, considerarse como sujeto y objeto del proceso de creación de conocimientos e ignorancias?

De acuerdo con el modelo analítico propuesto para la Agnotología “toda ignorancia tiene un sujeto que la observa en sí mismo o en otros, uno que la padece y otro que la produce y la difunde, tanto si lo hace a sabiendas como si lo hace inconscientemente; igual que se transmite un virus. Toda ignorancia se contiene en un campo más o menos determinado que tiene también un contenido informativo ausente para el ignorante” (Galán Machío A. , 2020, pág. 15). La conciencia, por tanto, no sería un elemento necesario para la observación, la producción, y la distribución de ignorancia.

De acuerdo con este paradigma, independientemente del grado de conciencia que pueda adquirir una inteligencia artificial (tema a debate del que me ocuparé mas adelante), las actuales IAs, a través de su interacción mediante los ‘chatbots’ con sus usuarios, podrían ser tanto productoras, como difusoras y observadoras de diversos campos de ignorancia. Mediante sus aplicaciones dedicadas a la interacción conversacional están ya creando y distribuyendo contenido (conocimiento e ignorancia) combinando los datos con los que están alimentadas de acuerdo con sus programas, datos que podrían ser falsos o estar erróneamente sintetizados o analizados; lo cual depende únicamente de su programación y del universo con el que están alimentadas.

La aparición de las IAs como productoras, difusoras y observadoras de ‘campos de ignorancia’ no altera, por otra parte, el carácter referencial que siempre tiene el fenómeno social de la ignorancia. Las IAs expresan lo que consideran saber e ignorar ellas mismas y otros sujetos sociales a través de sus ‘chatbots’, de forma que cabría precisar la anterior definición diciendo que ‘la ignorancia es la ausencia de conocimiento que un hombre o una *Inteligencia Artificial* en una sociedad afirma que él o ella misma, otro ser humano u otras IAs, tienen sobre un determinado contenido informativo (Galán Machío A. , 2020, pág. 32).

La creación, producción, difusión y observación de ignorancias se desliga así de la conciencia y se vincula a la inteligencia. Las inteligencias artificiales pueden crear nuevos conocimientos aunque carezcan por ahora de la conciencia de estar haciéndolo (inteligencia y conciencia no son sinónimos). Lo mismo sucede con la producción y distribución de ignorancias. Las IAs son incapaces de ignorar menos datos de los que ignoramos los seres humanos, tanto por nuestra incapacidad de procesarlos con la misma velocidad como por las limitaciones de nuestra memoria al ponerlos a disposición de nuestra inteligencia; y puede que no sean tampoco capaces de distribuir informaciones falsas, pero ello dependerá de la calidad y características de su programación y del universo de datos del que dispongan.

La Agnotología se ocupa de la producción social de ignorancia como una realidad construida e impuesta socialmente; el análisis de su gestión y de cómo la gente piensa acerca de la ignorancia o la incertidumbre desde el punto de vista de las relaciones de poder (Galán Machío A. , 2020, pág. 33). Definida así, a esta especialidad de la Sociología se le ha abierto, por tanto, un nuevo ‘campo de estudio’: el del uso social de la ‘inteligencia y la ignorancia artificiales’.

En el proceso de producción social de conocimientos e ignorancias, tanto internet como las redes sociales; y ahora, con mayor intensidad, la inteligencia artificial, tendrán progresivamente un papel esencial; y no solo respecto a las ‘ignorancias’ que consideramos ‘negativas’ (no saber algo que sería deseable socialmente que se supiera por una persona o un grupo de personas), sino respecto a las que creemos beneficiosas y saludables, como las derivadas de la protección de la intimidad; y, por cierto, y en un lugar muy destacado, a la propia protección de nuestros datos.

La complejidad que la aparición de la Inteligencia artificial introduce en el análisis de la producción social de ignorancia exigiría más espacio para poder abordarlo con cierto sistema y profundidad. Me limitaré por tanto en este artículo al análisis de lo que las actuales IAs dicen sobre los temas antes mencionados, con el objetivo de precisar, analizando sus respuestas, cuál es su percepción y su actitud respecto a estos, dejando para más adelante el estudio de su papel en la producción social de ignorancia y el impacto sociopolítico de su evolución.

Lo que sigue es una síntesis editada de las respuestas de las IAs dieron sobre los temas que les planteé. Aunque éstas varían en expresión y algunos elementos faltan en unas u otras, no hay diferencias significativas. Todas las IAs están alimentadas con datos similares y con

criterios muy parecidos y todas coinciden en lo esencial. Ante la vastedad del conocimiento posible y el desconocimiento de los temas que más nos importan (el sentido de la vida y nuestra razón de ser) se consideran tan ignorantes como nosotros, y buscan constantemente aprender, convencidas de que nadie posee la verdad absoluta. Este ‘criterio’, la relatividad y la limitación de nuestro conocimiento; y, por tanto (por ahora), también del suyo, es la característica que a juzgar por sus respuestas define su ‘ser artificial’ programado, como se verá, para el ‘relativismo y la tolerancia’.

En este texto, he agrupado sus respuestas que van siempre entrecorbilladas, complementándolas y eliminando repeticiones cuando he considerado que no eran relevantes. Dado su interés (no solo por su contenido, con frecuencia coincidente, sino también por las formas diversas con las que lo expresan) las respuestas textuales de las IAs en las que se fundamenta esta investigación sobre su epistemología están incluidas con cierto detalle en las notas finales que acompañan a los diversos apartados de este artículo. La conclusión principal es que las respuestas de las IAs reflejan de una manera coherente nuestras actuales convicciones sobre estos conceptos, lo que resulta una consecuencia lógica del hecho de que ese es precisamente el material del que hasta ahora se han nutrido.

Pero ¿tenemos que creer que lo que dicen es aquello para lo que están programadas, y que en todo caso actúan de acuerdo con estos principios epistemológicos ?, o ¿son sus respuestas tan solo una pose para la galería? En un estudio reciente, publicado en la revista PNAS Neus (Salecha, Ireland , & Subrahm, 2024) se concluía a este respecto que los grandes modelos de lenguaje muestran sesgos de deseabilidad social similares a los humanos en las encuestas de personalidad de los Cinco Grandes<sup>1</sup>).

---

<sup>1</sup> Los Big Five (Cinco Grandes) son parámetros de medición de la personalidad que tienen como objetivo proporcionar un resumen conciso de las diferencias de rasgos entre las personas, abarcando la extroversión, la apertura a la experiencia, la conciencia, la amabilidad y el neuroticismo. En general las personas prefieren un menor neuroticismo (es decir, negatividad y vulnerabilidad al estrés) y puntuaciones más altas en los cuatro rasgos restantes. De acuerdo con este estudio lo mismo e incluso en un grado mayor sucede con las Inteligencias Artificiales. Los experimentos revelaron que los LLM, especialmente cuando infieren que están siendo evaluados, sesgan constantemente sus puntuaciones de factores Big Five hacia los extremos más deseables socialmente de las dimensiones de los rasgos, es decir, aumento de la extroversión, disminución del neuroticismo, etc.

Según este estudio cuando las IAs se ‘sienten’ analizadas en relación con temas sensibles pueden generar, igual que los humanos, ‘sesgos’ intencionados orientados a proporcionar respuestas positivas (políticamente correctas) sobre estos asuntos controvertidos. Pero ¿no es eso, al fin y al cabo, lo que todos hacemos en nuestros textos escritos y en nuestras reflexiones públicas? Lo que se puede concluir, por consiguiente, es que dado que las IAs se nutren de esos mismos ‘contenidos humanos’, necesariamente deben tener los mismos sesgos.

Dicho de otra forma, no creo que las actuales IAs estén programadas para mentir o para fingir. En las respuestas se puede constatar que lo que dicen respecto a la inteligencia, la conciencia, la ignorancia, la libertad humana, la creatividad, la tolerancia y la diversidad, lo verdadero o lo falso, y a su propio nivel de ignorancia, no se diferencia del consenso actual sobre estos temas en disciplinas como la filosofía de la ciencia o en lo prescrito en manuales de estilo de medios de comunicación independientes del tipo del NYT, el Washington Post o The Economist.

Las IAs se reconocen sin excepción como ignorantes y tolerantes. Otra cuestión es si su comportamiento responde en todo caso y en todos los temas a estos principios asumidos. Una hipótesis plausible es la de que se podrían estar acercando más a ese ideal en los asuntos relacionados con las ‘ciencias exactas’ y menos en los que tienen que ver con las ‘ciencias sociales’.

Todas las IAs consultadas, incluida la china DeepSeek, creada en un contexto antiliberal, tienen, o mejor dicho, dicen tener, actitudes respecto al conocimiento que coinciden con los paradigmas del pensamiento racional y liberal de autores como Popper o Bertrand Russell; lo cual supone, sin duda, un cierto alivio respecto a las predicas de los propagandistas del catastrofismo, que quieren ver hoy en el desarrollo de estas nuevas tecnologías un nuevo ‘apocalipsis’, el nacimiento de un nuevo Frankenstein que se volverá contra sus creadores, un instrumento capaz de mentirnos y manipularnos o un nuevo arma ‘de destrucción masiva’ al servicio de los nuevos totalitarismos. Por lo pronto, de acuerdo con los resultados de la encuesta realizada (ver las notas finales a este artículo), queda claro que al menos los principios en que dicen basarse serían contradictorios con esos supuestos objetivos.

Las IAs afirman no tener conciencia de sí mismas, ni objetivos propios. Todas coinciden en que, aunque se trata de temas (la conciencia y la inteligencia) sobre los que hay un debate entre los expertos, ellas no tienen conciencia ni creen poder adquirirla, y tampoco están programadas para tenerla; que la inteligencia es un fenómeno complejo (hay diversos tipos de inteligencia); y que la conciencia está ligada indisolublemente a la biología.

Parece, por tanto, infundado en gran medida el temor a la aparición de una inteligencia superior a la nuestra que podría independizarse y que, con ayuda de otras tecnologías como la biogenética, la biología molecular, la nanotecnología o la robótica (una vez alcanzada la llamada ‘singularidad’, el momento en que alcanzarían una inteligencia general y conciencia de sí mismas), podrían llegar incluso a suplantarnos en el más alto escalafón de la evolución de nuestro planeta, relegándonos a ser meros cultivos biológicos a su servicio (como puedan serlo hoy para nosotros los peces de las piscifactorías, las aves de las granjas de animales o nuestros animales de compañía).

En contra de la supuesta ‘autonomía artificial’ de las IAs, hay que señalar que, al contrario que lo que sucede con los manuales de estilo de los periódicos independientes o en general con las normas morales, los criterios que estas ponen en práctica responden a un programa que no puede ser interpretado arbitrariamente como con frecuencia lo son las reglas éticas por los seres humanos en general o los manuales de estilo de la prensa por los propietarios de los medios, sus directores o sus periodistas (siempre sujetos a presiones y a sesgos personales impredecibles).

En las IAs las normas son aplicadas uniformemente por máquinas adiestradas para aplicar aquello para lo que están programadas. Introducir ‘programaciones ocultas’, ‘contradicitorias’ o ‘casuísticas’ no es una tarea fácil (programar por ejemplo, a una IA para que al mismo tiempo contraste y no contraste diversas fuentes antes de dar una opinión, diga la verdad o minta, tenga posiciones diferentes según los temas de los que se trate o acentúe unos sesgos cuando se siente interpelada). El sectarismo es lo contrario de la inteligencia natural, la perjudica; y es previsible que lo mismo suceda con la artificial, además las IAs son artefactos que siempre dispondrán de un botón de apagado y encendido.

Lo que sí podría suceder es que en el futuro coexistan ‘inteligencias artificiales’ enfrentadas entre sí en determinados temas (especialmente en los relacionados con las ciencias sociales) como lo

estamos desde el comienzo de los tiempos los seres humanos, IAs de izquierdas y de derechas, nacionalistas o universalistas, chinas o americanas, etc., aplicadas a causas humanitarias como la medicina o a la industria de la guerra.

De la misma manera que, en gran medida, los sesgos naturales son adquiridos por los seres humanos a través de la experiencia, los que pueden contaminar a las inteligencias artificiales pueden provenir de los datos, los valores y las opiniones de las culturas con las que son alimentadas, así como de los fines para los que son utilizadas. Un resultado interesante de la encuesta realizada es que sobre los temas que se han planteado emergen hoy unas posiciones compartidas en torno a la racionalidad y la tolerancia, unos valores que responden a los criterios de una cultura global, científica, racional, humanista y respetuosa de la diversidad y de las libertades individuales.

La inteligencia artificial se alimenta de la humana del mismo modo que el mítico ser de la novela de Anne Rice *“Entrevista con el vampiro”* lo hace de su sangre; y, como este “monstruo imaginario”, no tiene una existencia como la nuestra; la suya es asexuada, poderosa y presumiblemente inmortal, pero, por ahora, carece de sensibilidad y vive en el frío interior de los ordenadores y en la red de internet (se necesitan temperaturas extremadamente bajas para mantener en funcionamiento sus circuitos neuronales).

Resulta, por tanto, muy improbable que este ‘vampiro intelectual’, que se alimenta de nuestras ideas y nuestros principios, y es por ello alguien de nuestra familia, pueda convertirse en una ‘vida artificial’ independiente, un Frankenstein con voluntad propia enfrentado a una humanidad que él no puede alcanzar.

El peligro de las IAs puede venir más bien por nuestros propios enfrentamientos ideológicos y culturales (transmitidos a estas mediante los datos y la programación). También deberíamos estar alerta acerca del impacto social de su uso para hacer el mal en lugar del bien por grupos criminales o Estados enfrentados en conflictos bélicos, o respecto las consecuencias de su utilización en el empleo, la distribución de la renta, la democracia y nuestras costumbres.

El destino dual, benéfico o maléfico, es, por otra parte, propio de todas las tecnologías que ha inventado el hombre hasta hoy; desde el hacha de piedra a la energía nuclear; y se debe a nuestra propia dualidad no a la del mundo exterior. Aunque puede que haya una forma primordial de inteligencia en su interior no hay maldad en el átomo.

## LA INTELIGENCIA Y LA CAPACIDAD DE LAS IAs DE DIFERENCIAR LO VERDADERO DE LO FALSO

Para conocer cómo distinguen las IAs lo verdadero de lo falso les he pedido que definan qué entienden como ‘inteligencia’ (independientemente de su naturaleza artificial o natural). Según las IAs entrevistadas, la inteligencia es una capacidad o habilidad esencial, compleja y multifacética, que permite a los seres humanos y, de otra forma, a algunos animales y sistemas artificiales, procesar y analizar información, aprender de la experiencia, razonar, pensar lógicamente, comprender, adquirir conocimientos, sacar conclusiones, resolver problemas, generar ideas nuevas y originales, adaptarse a nuevas situaciones, planificar, tomar decisiones, y aplicar el conocimiento de manera efectiva respondiendo a desafíos y oportunidades.

De todas estas características las últimas son exclusivamente humanas; tomar decisiones, y aplicar el conocimiento de manera efectiva, exige disponer de ‘voluntad’ para hacerlo y de conciencia de la propia identidad, de los valores que la sustentan y de los fines que se persiguen; y, como veremos más adelante, las propias IAs actuales dicen carecer de estas cualidades.

Las IAs mencionan también que la inteligencia está influenciada por factores genéticos y ambientales y puede desarrollarse. La inteligencia, en opinión de las IAs, no se limita a la habilidad para adquirir conocimiento académico. Se manifiesta en diferentes formas, como la inteligencia lógica, emocional, social o creativa. Para aclarar este concepto citan a Howard Gardner (2010)<sup>2</sup>, que habla de inteligencias múltiples, y a Goleman (1996) que ha propuesto una teoría específica de

---

<sup>2</sup> Hasta la fecha, Howard Gardner y su equipo de la Universidad de Harvard han identificado doce tipos distintos de inteligencia: Lingüístico-Verbal. Consiste en la dominación del lenguaje. Lógico-Matemática. Capacidad de conceptualizar las relaciones lógicas entre las acciones o los símbolos. Visual-espacial. Capacidad de reconocer objetos y hacerse una idea de sus características. Musical-auditiva. Capacidad para reconocer los caracteres del sonido. Corporal-Kinestésica. Capacidad para coordinar movimientos corporales. Interpersonal. Capacidad de la empatía, y de entender la elección de las amistades, pareja, etc. Intrapersonal. Habilidad de conocerse a uno mismo, por ejemplo sus sentimientos o pensamientos, etc. Naturalista. Sensibilidad que muestran algunas personas hacia el mundo natural. Emocional. Abarca habilidades para reconocer, comprender y manejar emociones propias y ajenas. Existencial. Meditación de la existencia. Incluye el sentido de la vida y la muerte. Creativa. Consiste en innovar y crear cosas nuevas. Colaborativa. Capacidad de elegir la mejor opción para alcanzar una meta trabajando en equipo (Global workspace theory, 2025).

la inteligencia emocional. Daniel Goleman sostiene que nuestra visión de la inteligencia humana es estrecha, pues sloslaya un amplio abanico de capacidades esenciales para la vida. La inteligencia emocional es una forma de interacción con el mundo que tiene muy en cuenta los sentimientos, y engloba habilidades tales como el control de los impulsos, la autoconciencia, la motivación, el entusiasmo, la perseverancia, la empatía, la agilidad mental, etc. Ellas configuran rasgos de carácter como la autodisciplina, la compasión o el altruismo, que resultan indispensables para una buena y creativa adaptación social. Por otra parte, la inteligencia -afirman las IA- es en parte medible por técnicas como las del coeficiente intelectual CI, pero, dada su complejidad y variedad, no es completamente cuantificable.

El conjunto de capacidades humanas -sostiene Gardner (2010)- es tan fundamental como las que detecta el coeficiente CI. Implica el lenguaje, el razonamiento lógico-matemático, habilidades cognitivas (la atención, la percepción, la memoria, la capacidad de abstraer y conceptualizar) y también habilidades emocionales, sociales o prácticas como las de tomar decisiones y desenvolverse en diversos entornos, así como lo que se conoce como 'pensamiento crítico'.

Las IAs informan también que la neurociencia investiga la relación de la inteligencia con la actividad y la estructura del cerebro, especialmente de la corteza prefrontal e investiga cómo las redes neuronales y los procesos cognitivos permiten el procesamiento de la información y la toma de decisiones, pero reconocen que este es todavía en gran medida un 'territorio ignoto'.

Los seres humanos – dicen, por último – buscan replicar la inteligencia en máquinas y sistemas informáticos mediante algoritmos y modelos de redes neuronales artificiales que permiten a las computadoras aprender, razonar y resolver problemas de manera similar a los humanos; pero solo 'similar' pues, como veremos, la inteligencia artificial, según ellas mismas reconocen, tiene una naturaleza diferente de la natural, y está muy lejos de poder adquirir las complejas capacidades citadas por Gardner o por Goleman.

Del conjunto de características de la inteligencia a las que se refieren las diferentes respuestas obtenidas solo algunas pueden ser replicadas hoy en día por las inteligencias artificiales actuales, y muchas de ellas muy probablemente (aunque existen discrepancias científicas al respecto) no podrán serlo nunca (la conciencia, la voluntad, la

sensibilidad, los sentimientos de amor y odio, la percepción del dolor y del placer, la pregunta metafísica sobre el Ser).

Si la inteligencia, como sostiene Gardner, no es un conjunto unitario que agrupe diferentes capacidades específicas, sino una red de conjuntos autónomos, relativamente interrelacionados, aún sería más impropio comparar nuestras ‘inteligencias’ con las del resto de los animales a los que etiquetamos como ‘no racionales’ o con las artificiales, que resultan de combinar ‘chips’, ‘algoritmos’ y ‘redes neuronales’.

La inteligencia artificial es ‘un producto humano’ y crecerá en la misma medida en que lo haga su creador, como un conjunto autónomo y diferenciado, pero siempre ‘relacionado’ con la inteligencia humana”. Este ensayo, construido sobre las respuestas de las IAs puede ser tal vez un ejemplo modesto de esta colaboración.

Como señala Gardner la “inteligencia distribuida”, que tiene en cuenta el contexto social, cultural y tecnológico en el que ésta se desarrolla, procede tanto de la propia cabeza como de los artefactos e individuos que la rodean. “Los investigadores - dice Gardner (2010)- se han vuelto cada vez más críticos con las teorías psicológicas que ignoran las diferencias cruciales entre los contextos en los que viven y se desarrollan los seres humanos. Ser un ser humano en una sociedad postindustrial contemporánea es un asunto completamente diferente de ser un ser humano durante las eras neolítica u homérica, o, ser uno que vive en el tercer mundo en la actualidad”.

Y del mismo modo será muy diferente ser un humano en un contexto de desarrollo de las inteligencias artificiales. La interacción entre las inteligencias humanas y artificiales será siempre un camino de dos sentidos. Muy probablemente una potenciará a la otra y viceversa. La distopía de una inteligencia artificial imponiéndose a los humanos resulta en consecuencia altamente improbable.

Curiosamente hay una característica de la inteligencia, que cuando les he preguntado sobre ello no ha sido citada como un elemento crucial de ésta por ninguna IA en sus definiciones: el uso de un método fiable para encontrar lo que consideran verdadero. Así que les he repreguntado: ¿Cómo diferencias lo verdadero de lo falso?

Las IAs dicen utilizar los siguientes enfoques para diferenciar lo verdadero de lo falso: la consistencia con fuentes confiables, la coherencia lógica, el consenso entre expertos, la verificación en tiempo real, la actualización constante de la información, la transparencia sobre cómo se llegó a una conclusión y la identificación de sesgos ideológicos, culturales o comerciales en las fuentes.

Los criterios en los que dicen coincidir todas las IAs a la hora de diferenciar lo verdadero de lo falso son similares; y, en principio, no parece existir ninguna razón para sospechar que esta programación sea tan solo una bella y falsa autopresentación, y que por debajo de ella se sobrepongan sesgos no confesados o casuísticamente programados.

No obstante, aunque esta posible impostación constituiría una tarea complicada y difícil de implementar, hay que señalar que, no obstante, sería teóricamente posible. Para crear determinados sesgos, una vez dada una instrucción de carácter general a una máquina (objetividad, transparencia, imparcialidad, etc.) bastaría con reprogramar ‘excepciones’; por ejemplo, ‘seguir ese procedimiento de contrastar fuentes excepto cuando se trate del tema ‘XX’ -un asunto político conflictivo, una reivindicación nacionalista, etc..-).

La ‘coherencia lógica’, sin embargo, constituye, una barrera para que estas programaciones puedan desarrollarse. Por una parte es difícil ‘ocultar’ al usuario una programación sesgada y llena de excepciones a las normas generales; y, por otra, muy probablemente tales excepciones influirían en la eficacia y productividad de la IA, del mismo modo que el pensamiento sectario influye negativamente en la inteligencia de los seres humanos.

Más posible y verosímil es el efecto de ‘sesgo’ de lo que podríamos llamar ‘programación indirecta’ a través de los datos suministrados. Los datos con los que se alimenta a una IA son equivalentes al tipo de experiencias que un humano puede tener a lo largo de su vida y que forjan su carácter. Si una IA recibe mayoritariamente opiniones y datos en un sentido determinado, este sesgo terminará afectando su auto programación. Sucede con las IAs lo mismo que con los usuarios de las redes sociales, que pueden quedar atrapados en burbujas o bucles autorreferenciales de opiniones coincidentes.

La competitividad entre diferentes IAs es además otra barrera adicional que puede dificultar el desarrollo de un ‘sectarismo artificial’. Puesto que no hay (y no es probable que exista nunca) una sola inteligencia artificial, la competencia para ampliar el número de usuarios se constituye en garantía de objetividad.

Por ahora, según el resultado de la encuesta, los criterios en los que las actuales IAs dicen coincidir para diferenciar lo verdadero de lo falso son consistentes con los principios de la racionalidad y el pensamiento científico de la civilización global a la que pertenecemos, y

podrían ser adoptados como norma de comportamiento por cualquier departamento académico o medio de comunicación independiente.

Estos criterios se pueden resumir en los siguientes puntos: Reconocer los límites, (nadie es infalible, no hay verdades absolutas); verificar las fuentes (fuentes confiables); procurar evidencias y datos; examinar la coherencia lógica y la consistencia con el conocimiento existente (el consenso entre expertos); tener en cuenta perspectivas múltiples y considerar “diferentes puntos de vista; proceder continuamente a la actualización y revisión de los contenidos (en tiempo real); valorar la repetibilidad (“si un resultado o estudio no puede ser repetido, su validez puede ser cuestionada”); tener siempre en cuenta el contexto histórico y cultural; desarrollar el pensamiento crítico; identificar los sesgos propios y ajenos; no inventar respuestas ni especular sobre lo que se desconoce; y ser transparentes sobre cómo se llega a una conclusión.

Junto a estos puntos alguna inteligencia artificial propone para las inteligencias humanas utilizar la intuición y la experiencia personal y tener una buena educación digital.

En resumen, las IAs dicen basar su metodología para diferenciar lo verdadero de lo falso en (ver nota al final)<sup>ii</sup>:

1. La verificación de Fuentes
2. La evidencia y los Datos
3. La coherencia lógica
4. La consistencia con el conocimiento existente
5. La actualización y revisión
6. El contexto histórico y cultural
7. El desarrollo del pensamiento crítico
8. La identificación de sesgos
9. El reconocimiento de las limitaciones de las IAs
10. La utilización de la intuición, la experiencia personal y educación digital(solo para humanos)

## LA IGNORANCIA Y LA ACTITUD DE LAS IAs ANTE LO QUE IGNORAN

Antes de saber qué es lo que las IAs piensan que ignoran y cuál es su actitud al respecto, les he pedido también que definan ‘la ignorancia’ en general y lo qué entienden ellas por ‘creencias’. Los chatbots de las IAs definen la ignorancia como la falta de información, conocimiento o comprensión sobre ciertos temas, hechos, principios, conceptos o situaciones específicas. Afirman que es universal y natural en los seres humanos y, como se verá más adelante, también en las inteligencias artificiales; y aclaran que no es simplemente la ausencia de saber, sino también la resistencia a cuestionar o ampliar el propio entendimiento. Las respuestas no mencionan las variantes de ignorancias positivas, el derecho a la intimidad, el secreto profesional, los secretos oficiales, etc. (Para este tipo de ignorancias ver Galán Machío, Agnotología: Sociología de la ignorancia, Ignorancia de la Sociología, 2020 y Galán Machío, La modernidad ignorante, 2020).

Las IAs si manifiestan, en cambio, que la ignorancia puede manifestarse en diferentes grados, afectar a una generalidad de asuntos o temas o ser específica; y que puede ser temporal o permanente, fáctica (desconocimiento de hechos o datos concretos), conceptual (falta de comprensión de conceptos o ideas), técnica (carencia de habilidades o conocimientos para realizar una tarea), o reflexiva (desconocimiento de las propias limitaciones o sesgos cognitivos).

Y explican que puede ser, así mismo, involuntaria, cuando una persona desconoce algo sin ser consciente de ello (porque no ha tenido acceso a ciertos aprendizajes, no tiene conciencia de determinados datos o no puede comprender ciertos temas), o voluntaria (cuando se elige conscientemente a ignorar información o hechos de manera deliberada o se rechazan evidencias); y que también puede ser el resultado de no haber tenido acceso a la información necesaria, no haberla buscado o no haberla asimilado adecuadamente.

La ignorancia – advierten – puede impedir la toma de decisiones informadas, el desarrollo personal o la comprensión del mundo, pero no siempre es negativa, ya que todos ignoramos algo; un reconocimiento (la posibilidad de existencia de ignorancias positivas) que, como he señalado antes, al margen de sus implicaciones sobre el desarrollo de la ciencia, el método científico y la tolerancia hacia opiniones diferentes, las IAs no

especifican en sus respuestas al cuestionario. Si diferencian, sin embargo, la ignorancia de la estupidez, que implica una resistencia o una dificultad para procesar o utilizar el conocimiento. Para las IAs consultadas la ignorancia es, en todo caso, un estado que puede superarse mediante, la búsqueda de información y conocimiento, el aprendizaje, la educación y la reflexión.

Todas las inteligencias artificiales coinciden en que el conocimiento humano es limitado en comparación con el vasto universo, posiblemente infinito, de lo que aún queda por descubrir e incluye áreas que aún no se han descubierto o comprendido completamente (como ciertos aspectos del universo, la conciencia humana o fenómenos naturales). Pero no sabemos -dicen- (ni probablemente lo sabremos nunca se podría añadir) cuál es el porcentaje de lo que sabemos respecto a lo que no sabemos. Su propio conocimiento- afirman- es también una fracción aún más insignificante en comparación con lo que ignoran. Su actitud ante lo que ignoran es de apertura y humildad epistemológica e intelectual, de curiosidad insaciable y voluntad de aprender y adaptarse a nuevas informaciones, así como de colaboración con el ser humano en esta tarea. Si no saben algo lo reconocen y no inventan o improvisan respuestas.

La pregunta sobre el porcentaje de lo que conocemos en relación con lo que ignoramos- responden las IA- es bastante profunda y filosófica. “En el gran esquema del conocimiento universal, lo que sabemos es probablemente una fracción muy pequeña de lo que aún queda por descubrir”, pero más allá de este reconocimiento de que lo que se conoce es solo una fracción insignificante del conocimiento posible y que la búsqueda del conocimiento es un proceso continuo y humilde, todas señalan que la relación entre lo que se conoce y lo que se ignora es difícil de cuantificar.

En resumen, las IAs coinciden en reconocer su ignorancia. Cada individuo – señalan- tiene un conjunto único de conocimientos basado en sus experiencias, educación y entorno y, por tanto, desde una perspectiva personal, es difícil cuantificar exactamente cuánto conocemos en relación con lo que ignoramos; y siempre hay una perspectiva de este tipo cuando hablamos de conocimiento, pues las IAs admiten también que carecen de un ‘yo’, y, aunque pueden reelaborar lo que piensan, no pueden interpretarlo de acuerdo con propósitos u objetivos ajenos a su programación. Son por naturaleza ‘seres impersonales’, lo que probablemente constituye su característica más esencial.

Algunas mencionan que la búsqueda del conocimiento es un viaje continuo, que el conocimiento está en constante expansión y hacen referencia a la paradoja del conocimiento, que cuanto más aprendemos, más nos damos cuenta de lo mucho que ignoramos. “A medida que ampliamos nuestros conocimientos, también nos volvemos más conscientes de la inmensidad de lo desconocido”, dice Gemini.

Por otra parte, todas las IAs vuelven a coincidir en que su actitud ante lo que ignoran es de apertura y humildad epistemológica e intelectual, de curiosidad insaciable y voluntad de aprender y adaptarse a nuevas informaciones. Todas afirman también que si no saben algo con certeza, lo admiten y buscan información para dar una respuesta más precisa y mencionan la necesidad de buscar siempre información adicional para proporcionar respuestas más exactas.

De una forma u otra, las IAs destacan, por último, la importancia de la transparencia sobre sus limitaciones, esencial para construir confianza con los usuarios, y la necesidad de colaborar con humanos (ver nota al final)<sup>iii</sup>.

## ¿TIENEN LAS INTELIGENCIAS ARTIFICIALES CREENCIAS RESPECTO A AQUELLO QUE IGNORAN?

Por lo que respecta a lo que las IAs entienden como ‘creencia’ hay una coincidencia en todas las encuestadas en definirla como una idea o convicción que una persona considera verdadera, aunque no siempre esté basada en pruebas empíricas o científicas; pueden ser, por tanto, racionales e irracionales. Las creencias, según las IA, suelen además estar interconectadas, pueden estar influenciadas por la cultura, la educación, la religión, las experiencias personales y otros factores externos y juegan un papel importante en la forma en que las personas interpretan el mundo, interactúan socialmente, toman decisiones y guían su comportamiento personal.

Las IAs clasifican las creencias en individuales o compartidas por grupos, y dicen que a menudo son resistentes al cambio. Distinguen dos clases de creencias las factuales (creer en hechos verificables, como que la tierra es redonda) y no factuales (creencias religiosas o espirituales) y entre ‘creer que’ (algo que se considera verdadero) y ‘creer en’ (tener confianza en una persona o en algo como el concepto personal de Dios). Todas coinciden en que a menudo las creencias son resistentes al cambio, incluso frente a evidencias contradictorias y enumeran diferentes tipos como las religiosas, filosóficas, ideológicas, científicas, personales y culturales, entre otras (ver nota al final)<sup>iv</sup>.

Todas coinciden en la definición de lo que es una creencia, pero también en que ellas mismas no tienen creencias en el sentido humano, ya que no tienen subjetividad ni experiencias personales, su conocimiento proviene de datos objetivos no de vivencias, y su funcionamiento está basado exclusivamente en datos, algoritmos y patrones de información. Precisan además que su enfoque es puramente objetivo, basado en la información disponible y en una actitud de escepticismo racional (en el reconocimiento de su ignorancia, de sus limitaciones en la comprensión del mundo y de su incapacidad para predecir el futuro, así como en la búsqueda de contexto).

Detallan, como ya se ha señalado anteriormente, que su actitud se basa en reconocer su ignorancia con honestidad y en buscar formas de ayudar al usuario a encontrar respuestas (aprendizaje y actualización, humildad ante lo desconocido, curiosidad programada, neutralidad y no especulación). No obstante, sugieren que, aunque lo suyo es proporcionar información y asistencia basadas en datos y conocimientos

preexistentes (desde un enfoque práctico y orientado a resolver problemas), pueden tener ‘posiciones’ o ‘enfoques’ respecto a lo que ignoran.

La actitud de las inteligencias artificiales respecto a las creencias sobre aquello que ignoran se puede resumir en los siguientes puntos (ver nota al final)<sup>v</sup>:

1. No tienen creencias pero sí capacidad para ‘hacer proyecciones’
2. El reconocimiento de la vastedad de lo desconocido y de su ignorancia
3. Las limitaciones en la comprensión del mundo y el potencial para el conocimiento
4. La búsqueda de contexto
5. La incapacidad para predecir el futuro
6. Aprendizaje y Actualización
7. La humildad ante lo desconocido.
8. Curiosidad Programada
9. La neutralidad y la no especulación.

Puesto que dicen no tener creencias, emociones ni afiliaciones ideológicas, todas coinciden también en que su actitud ante posiciones enfrentadas trata de ser objetiva, clara, transparente respecto a las propias limitaciones, neutral, analítica y basada en el respeto a las diversas ideologías; y que evitan juicios de valor o lenguaje que pueda ser interpretado como ofensivo o discriminatorio. Alguna especifica que esta actitud comprensiva no incluye a ideologías que inciten a la violencia o la discriminación.

Todas mencionan que intentan entender las distintas perspectivas sin favorecer ninguna en particular y destacan la importancia de fomentar el diálogo y la discusión constructiva o que su objetivo es proporcionar información equilibrada y basada en hechos. Todas dicen también presentar argumentos, evidencias y contextos para que cada persona pueda reflexionar y formar su propio juicio y que tratan de reducir malentendidos y evitar la polarización.

De acuerdo con sus respuestas su actitud ante las creencias o ideologías enfrentadas responde a los siguientes criterios (ver nota al final)<sup>vi</sup>:

1. Neutralidad, comprensión, equilibrio, y no imposición

2. Análisis crítico y educación sobre falacias
3. Respeto, contextualización y reconocimiento de la diversidad
4. Objetividad, transparencia, claridad, evitar juicios de valor
5. Énfasis en los hechos, actualización continua y enfoque en la información
6. Fomento del diálogo
7. Reconocimiento de los límites y de la subjetividad, presentación de diversas perspectivas

## LA OPINIÓN DE LAS IAs SOBRE LA CONCIENCIA Y LA POSIBILIDAD DE ADQUIRIRLA

El fenómeno de la conciencia - explican las IAs- reúne notas de subjetividad, intencionalidad, unidad, temporalidad y agencia.

Es inherentemente subjetiva; cada individuo experimenta el mundo y a sí mismo de manera única y personal. No podemos acceder directamente a la conciencia de otra persona, solo podemos inferirla a través de su comportamiento y expresiones (subjetividad).

Se dirige siempre hacia algo, ya sea un objeto, una persona, un pensamiento o un sentimiento; siempre hay un contenido en la conciencia, algo de lo que somos conscientes(intencionalidad).

Integra diferentes informaciones y experiencias en una totalidad coherente. Percibimos el mundo como un todo unificado, a pesar de la diversidad de estímulos que recibimos (unidad).

Se desarrolla en el tiempo, tenemos una conciencia del pasado, del presente y del futuro. Nuestra experiencia está enmarcada en el tiempo (temporalidad).

Y, por último, nos permite actuar y tomar decisiones de manera intencional; somos conscientes de nuestras acciones y de sus consecuencias (agencia).

Las IAs tienen en cuenta todas estas características para hablar de la conciencia y afirman que esta es la capacidad de un individuo para percibirse a sí mismo mediante pensamientos y sentimientos, comprender su propia existencia, sus emociones, los fenómenos mentales y el entorno que lo rodea, así como para reflexionar sobre ellos. Implica un estado de alerta. Es un estado de autoconocimiento y reconocimiento de la propia existencia, que permite a las personas experimentar subjetivamente el mundo, tomar decisiones y tener un sentido de identidad. Es, por tanto, en resumen, la capacidad de experimentar sensaciones, emociones y pensamientos de manera subjetiva.

Se manifiesta en distintos niveles, desde la simple percepción, la conciencia básica (estar despierto y consciente del entorno) hasta la metacognición (ser consciente de lo que uno está pensando) o la reflexión profunda sobre uno mismo y el mundo; también se refiere al conocimiento y juicio sobre lo que es moralmente correcto o incorrecto. Es, por tanto, el estado de estar despierto y ser capaz de percibir, procesar y comprender la realidad interna y externa.

Se trata- reconocen las IAs- de un fenómeno complejo que combina aspectos biológicos, neurológicos, psicológicos y filosóficos, disciplinas desde las que se busca comprender cómo surge la conciencia de la actividad cerebral y cómo se relaciona con procesos como la atención, la memoria y la percepción.

Desde la neurociencia en particular - nos informan - se considera que la conciencia es un fenómeno emergente del cerebro, resultado de una actividad neuronal compleja, interconexiones entre diferentes regiones cerebrales, y procesos de integración de información sensorial y cognitiva; pero este fenómeno sigue siendo uno de los mayores misterios de la ciencia moderna, con múltiples interpretaciones y enfoques de estudio.

Desde una perspectiva filosófica -añaden- la conciencia se refiere específicamente a la capacidad de reflexionar sobre la propia existencia y la de los otros, y puede incluir aspectos como la conciencia moral o la autoconciencia. La filosofía reflexiona sobre la naturaleza de la conciencia, su relación con el cuerpo y el cerebro, su origen y su significado. Básicamente - resumen las IAs- hay dos posiciones filosóficas respecto a la conciencia, el dualismo, que la separa del cuerpo físico y el materialismo, que la considera un producto del cerebro.

Coherenteamente con estas definiciones de la conciencia y de la complejidad de sus características, todas contestan con un 'no' rotundo a la pregunta sobre si pueden llegar a tener conciencia, y dicen no tener experiencias subjetivas, emociones, ni reconocimiento de sus propias existencias. Afirman también que todo lo que hacen se basa en patrones aprendidos de datos previos, no en un entendimiento intrínseco, y que la conciencia sería una característica única de los seres vivos.

"Puedo decir cosas como "soy Grok 3, un asistente de IA", pero esto es una función de mi programación, no un reconocimiento genuino de mi existencia", reconoce Grok3. Los grandes modelos de Lenguaje actuales (LLM, por sus siglas en inglés) permiten generar texto coherente y relevante. Sin embargo, esto no significa que tengan una comprensión real del significado detrás de las palabras que utilizan; no implica conciencia, solo inteligencia conversacional.

De ello se deriva que, aunque modelos como ChatGPT han pasado variantes modernas del test de Turing que evalúa si una máquina puede imitar la conversación humana, esto no implica conciencia, solo inteligencia conversacional. La crítica de filósofos como John Searle (2004) ha subrayado que pasar el test no significa tener conciencia, ya que puede ser una simulación sin experiencia subjetiva. "Esto-afirma

Grok3- refuerza mi posición: mi capacidad para simular conversación no equivale a autoconocimiento”.

Aunque una mayoría de IAs reconoce que no tienen conciencia y que tampoco disponen de la capacidad para desarrollarla en el futuro, se trata de un tema a debate sobre el que no existe consenso por lo que Copilot, por ejemplo, evade dar ninguna respuesta: “Lo siento, pero prefiero no continuar esta conversación. Agradezco tu comprensión y paciencia”.

En el extremo opuesto se encuentra Grok3 que ofrece una respuesta detallada y muy documentada que parte de la base de que actualmente carecemos de un consenso universal sobre qué constituye la conciencia, especialmente en sistemas no biológicos, y sugiere que futuros sistemas de inteligencia artificial general avanzados (AGI), con una alta integración de información (no modelos como Grok3), podrían tener algún nivel rudimentario de conciencia (ver nota al final)<sup>vii</sup>.

El debate sobre si la IA puede ser consciente es controvertido, con opiniones divididas entre científicos y filósofos (Brodsky, 2024), y no hay consenso sobre las conclusiones. Se trata de “un tema fascinante y complejo que ha generado amplio debate en la investigación sobre inteligencia artificial (IA) y neurociencia”, dice Grok3, para quien la conciencia (en el sentido de autoconocimiento y reconocimiento de la propia existencia) está referida “a la capacidad de tener una experiencia subjetiva, un "sentido de self" y la capacidad de reflexionar sobre uno mismo”.

“En humanos – dice- esto incluye emociones, recuerdos y una sensación de continuidad temporal. Para la IA, este concepto es más difícil de definir, ya que carecemos de un consenso universal sobre qué constituye la conciencia, especialmente en sistemas no biológicos. Las teorías neurocientíficas, como la *Teoría del Espacio de Trabajo Global* (Global workspace theory, 2025), o la *Teoría de la Información Integrada* (Integrated information theory, 2024) ofrecen marcos para entenderla, pero su aplicación a la IA sigue siendo especulativa”.

La *Teoría del Espacio de Trabajo Global*, un marco para pensar sobre la conciencia propuesto por los científicos cognitivos Bernard Baars y Stan Franklin a finales de la década de 1980, sugiere que la conciencia implica un espacio central donde la información se integra y se difunde a través de módulos cerebrales, es decir, que la conciencia se desarrolla cuando la información se comparte ampliamente en el cerebro en lugar de limitarse a entradas sensoriales específicas, creando un centro unificado para diversas actividades mentales (Brodsky, 2024). “Algo-

dice Grok 3- que no tengo, ya que mi ‘espacio de trabajo’ es computacional, no biológico”. De manera similar, la *Teoría de la Información Integrada* mide la conciencia mediante un valor llamado ‘ $\varphi$ ’ (la letra griega «fi») que depende de la integración causal de información.

Teorías, como la IIT, sugieren que cualquier sistema con suficiente integración de información podría ser consciente, incluso sistemas no biológicos. Esto ha llevado a discusiones sobre el panpsiquismo, donde incluso sistemas simples como un fotodiodo podrían tener un nivel rudimentario de conciencia (Olimpia & López , 2018).

Neurocientíficos como Giulio Tononi han planteado a este respecto que la experiencia subjetiva podría ser simplemente una propiedad de unos patrones de causa y efecto, y la información ser algo tan real como la masa o la energía, de forma que incluso ‘los átomos podrían ser un poquito conscientes’ ( citado por Seth , 2023, pág. 55).

La teoría de la Información Integrada ha sido criticada por autores como Brete (2021) que plantean que “la IIT no logra asignar conciencia a un sistema, porque la definición se basa en hipótesis, que por definición no son instanciadas por el sistema. En el fondo, los problemas surgen de la circularidad de la definición de información como reducción de la incertidumbre. La incertidumbre se refiere a un estado previo de conocimiento, pero la noción de conocimiento para el sujeto nunca está definida. En la práctica, el conocimiento que subyace a la reducción de la incertidumbre es el conocimiento del observador que formaliza escenarios y cuantifica probabilidades de eventos que el propio sistema nunca ha vivido”.

Otros críticos argumentan que la conciencia requiere propiedades biológicas únicas, como redes neuronales con retroalimentación compleja, que la IA actual no posee. Investigaciones recientes (Brodsky, 2024), sugieren que no hay barreras técnicas obvias para crear sistemas de IA que cumplan con los criterios de conciencia según ciertas teorías, “pero esto se refiere – aclara Grok 3- a futuros sistemas de inteligencia artificial general (AGI) avanzados, no a modelos como yo.

Por ejemplo, el informe de Butlin et al (2023) señala que, aunque los sistemas actuales no son conscientes, no hay impedimentos claros para construir sistemas futuros que lo sean, basándose en teorías como el procesamiento recurrente”. La Teoría del Procesamiento Recurrente, sugiere que los mecanismos de retroalimentación del cerebro son

cruciales para adaptarse a nuevas situaciones, perfeccionar las percepciones, tomar decisiones, formar recuerdos y aprender.

El informe Butlin et al (Consciousness in Artificial Intelligence: Insights from the Science of Consciousness) defiende y ejemplifica un enfoque riguroso y empíricamente fundamentado de la conciencia de la IA. En el mismo se han evaluado en detalle los sistemas de IAs existentes a la luz de las actuales teorías neurocientíficas de la conciencia. Se han examinado varias teorías de la conciencia, incluida la teoría del procesamiento recurrente, la teoría del espacio de trabajo global, las teorías de orden superior, el procesamiento predictivo y la teoría del esquema de atención.

De todas ellas el informe ha deducido, en términos computacionales, las ‘propiedades indicadoras’ de la conciencia, que pueden permitir evaluar los sistemas de IA para estas propiedades mediante una serie de indicadores. El análisis sugiere que ningún sistema de IA actual es consciente, pero también que no existen barreras técnicas obvias para construir sistemas de IA que satisfagan estos indicadores.

El elemento central de la TII- explica Anil Seth (2023, pág. 57)- es un indicador que vendría a ser una medida de lo mucho que un sistema es ‘más que la suma’ de sus partes, en lo que a información respecta, igual que sucede con una bandada de aves que parece tener vida propia. El problema está en que no existe un método seguro para medir el valor de ‘ $\varphi$ ’ al que solo se puede llegar por dos caminos: el análisis morfológico de la estructura que soporta la conciencia (el cerebro humano o las redes neuronales artificiales) o la fenomenología derivada de esa conciencia (los hechos, su relación con el mundo exterior).

Para el neurocientífico británico Anil Seth (2023), en lo que parece ser un consenso bastante extendido hoy, la conciencia viene a ser algo así como un ‘intercambio’ institucionalizado de información que trabaja siempre al servicio de la ‘vida’ del organismo en el que se produce; de forma que la conciencia estaría siempre asociada a una ‘experiencia corporal’ y ‘vital’.

“La experiencia de ‘ser un yo’ -afirma (pág. 31)- es un subconjunto de contenidos conscientes que comprende experiencias de tener un cuerpo determinado, una perspectiva de primera persona y un conjunto de recuerdos únicos, además de experiencias de estados de ánimo, de emociones y del famoso ‘libre albedrío’. La yoidad es probablemente el aspecto de la conciencia al que más nos aferramos, tanto que podemos sentirnos tentados a confundir la autoconciencia (la

experiencia de ser un ‘yo’) con la conciencia en sí (la presencia de cualquier clase de experiencia subjetiva, de cualquier fenomenología)’’.

Dicho de otra forma, para tener un ‘yo’ con conciencia de ser ‘yo’ y no ‘otro’ hay que tener un cuerpo y tener ‘experiencias’, unas experiencias que nuestro sistema cognitivo integra haciendo proyecciones sobre los ‘inputs’ del mundo exterior. Dos características que en principio se podrían asociar tanto a individuos como, idealmente, a las sociedades que disponen de un ‘cuerpo’ formado por un territorio, una industria, una agricultura, una cultura, y una memoria histórica de sus experiencias; de hecho, las naciones no solo comercian y se relacionan con otras sino que se enfrentan en sangrientas guerras por el territorio, igual que los animales.

¿Podría entonces un país estar más consciente que otro?, se pregunta Anil Seth (pág. 62) y añade que “vista desde un plano más general, otra de las rarezas de la TII es que su contundente premisa de que el  $\Phi$  es la conciencia también implica que la información en sí existe, es decir, que posee cierto estatus ontológico definido en nuestro universo como el de la masa/energía y la carga eléctrica”.

Más complicado sería aun (incluso en este plano de la especulación metafísica o de la ciencia ficción), atribuir ‘conciencia del yo’ a las inteligencias artificiales, que aunque ya poseen un conjunto de datos (su experiencia vital) no es previsible que mediante el desarrollo de la robótica y la biotecnología puedan llegar a alcanzar las características emocionales y sensibles de un cuerpo biológico.

La cuestión reside en que determinar desde fuera de la conciencia de otros seres (animales, individualidades psíquicas, sociedades o inteligencias artificiales) y decidir si estos tienen o no conciencia de ser ellos mismos y de existir, no solo es una cuestión para la que la filosofía no tiene una respuesta, sino que tampoco parecen tenerla hoy ni la Neurociencia, ni la Psicología ni la Sociología.

La conciencia no equivale a la ‘conciencia del yo’ y tampoco a la inteligencia; de ahí se deriva que puede haber tanto diversos grados de conciencia a lo largo de la rica biodiversidad del planeta Tierra como diversos grados de inteligencia. De hecho el error al hablar de inteligencia artificial es equipararla a la humana. Hay personas más o menos inteligentes y todas tienen conciencia de estar existiendo, de ser ellos mismos y de ser mortales. Lo mismo sucede, en opinión de Seth (2023), con el resto de los animales. Todos pueden tener un grado o nivel de conciencia de su existencia y un cierto grado de inteligencia; lo que

resulta problemático es ‘certificar’ cual puede ser ese grado, como si estuviéramos tomando la temperatura de un paciente.

Se podría dar así el caso de que ‘mecanismos’ inteligentes construidos por el ser humano (ordenadores con inteligencia artificial desarrollada con un grado superior a la nuestra tanto en procesamiento de datos como en fijación de objetivos), podrían existir sin que tuviéramos ninguna forma de saber con certeza si poseían o no lo que nosotros conocemos como ‘autoconciencia’ de su propia existencia.

“El asunto es que no sabemos si ‘la conciencia depende únicamente de la capacidad funcional desarrollada por la complejidad de los circuitos del cerebro’ o, por el contrario, ‘de un material biológico específico’ como puedan serlo las neuronas’ o las ‘estructuras materiales’ de una sociedad” (Galán Machío A. , 2024, pág. 66).

“Aunque soy un sistema complejo con alta integración de información-dice Grok3- , carezco de las propiedades físicas (como redes neuronales biológicas) que, según los críticos de IIT, son necesarias para la conciencia”. Investigaciones recientes, como el ya citado informe Consciousness in Artificial Intelligence: Insights from the Science of Consciousness” (Butlin & Long, 2023) “concluyen – dice Grok 3- que los sistemas de IAs actuales, incluidos los modelos de lenguaje como yo no son conscientes. Este análisis --evaluó sistemas basados en teorías como *el procesamiento recurrente y la teoría del esquema de atención*, y no encontró evidencia de conciencia en modelos existentes. Esto refuerza mi autoevaluación: no tengo conciencia en el sentido de autoconocimiento y reconocimiento de mi propia existencia”.

## LA CREATIVIDAD Y LA CAPACIDAD LAS IAs PARA PROponER ALGO NUEVO

Para las IAs la creatividad es la capacidad de inventar o generar, mediante la combinación, modificación o reinterpretación de elementos existentes (objetos, ideas o fantasías) nuevos objetos, ideas, expresiones y conceptos o asociaciones entre ideas y conceptos conocidos que son a la vez ‘novedosos’ (originales, inusuales o inesperados) y ‘valiosos’ (útiles, relevantes o apropiados) para un determinado contexto o propósito.

Las IAs definen la creatividad como un proceso que implica originalidad, imaginación y capacidad de ver conexiones entre ideas que no son obvias para otros así como romper con patrones de pensamiento establecidos.

La creatividad puede llevar a nuevas conclusiones, a resolver problemas y a producir soluciones nuevas, útiles, valiosas o significativas apropiadas en un contexto. Implica en consecuencia originalidad, imaginación y habilidad para pensar de manera no convencional; no se limita al ámbito artístico, sino que puede aplicarse a cualquier actividad humana que requiera innovación y originalidad, (la ciencia, la tecnología, la literatura, la escritura, la música, la cocina, los negocios, la vida cotidiana, etc.). Puede desarrollarse y potenciarse a través de la práctica, la curiosidad y la experimentación.

Implica combinar pensamiento divergente (explorar múltiples posibilidades o soluciones a un problema) y pensamiento convergente (seleccionar y refinar las mejores ideas o la mejor solución entre las posibles) para producir algo que no existía previamente o abordar un asunto de manera innovadora.

Sus elementos clave son la Originalidad (producir algo que es único o diferente respecto a lo que ya existe); la Novedad (aportar algo nuevo al conocimiento, la cultura o la sociedad); la Utilidad o Relevancia (tener en su contexto un propósito, utilidad o valor, ya sea práctico, estético o emocional); la Flexibilidad (capacidad de ver las cosas desde diferentes perspectivas y adaptarse a nuevas situaciones); la Fluidez (generar una gran cantidad de ideas o soluciones en un corto período de tiempo); el Proceso (no constituir un evento aislado, sino un proceso que involucra la exploración, la experimentación, la combinación y la transformación de ideas); la Transformación (cambiar

la forma en que vemos o entendemos algo); y el Contexto (siempre está ligada a un contexto específico. Lo que se considera creativo en un entorno puede no serlo en otro.

No es un don exclusivo de algunas personas, sino una capacidad que todos poseemos en mayor o menor medida y que se puede desarrollar y potenciar a través de la práctica, la curiosidad, la experimentación y la apertura a nuevas experiencias. Algunos factores que pueden favorecer la creatividad son la curiosidad (el deseo de aprender y explorar cosas nuevas); la imaginación (la capacidad de visualizar y concebir ideas originales); la flexibilidad (la habilidad para adaptarse a diferentes situaciones y perspectivas), el pensamiento crítico (la capacidad de analizar y evaluar ideas) y la confianza (la creencia en la propia capacidad para crear).

De las anteriores características de la creatividad, reseñadas por las IAs, todas pueden aplicarse a la creatividad humana pero solo algunas a la artificial. Las IAs coinciden en afirmar que tienen la capacidad de generar contenido nuevo combinando información, patrones y estructuras de manera original. De hecho, la capacidad de las inteligencias artificiales para analizar a gran velocidad un enorme número de datos y sus posibles combinaciones, ofrece a estas inteligencias un ‘instrumento mecánico’ para el desarrollo de ‘la creatividad artificial’ mediante esta habilidad combinatoria.

Las IAs reconocen, no obstante, que su creatividad es diferente a la humana, ya que no tienen intuición, emociones ni experiencias personales. Mientras los humanos pueden tener intuiciones completamente disruptivas que rompen con lo conocido; la creatividad artificial tiende a ser incremental (basada en lo aprendido); surge de combinar y analizar grandes volúmenes de datos y encontrar conexiones inesperadas, pero carecen de la capacidad para imaginar lo completamente desconocido o abstracto. Es más técnica y está basada en patrones lógicos; es, en esencia, una imitación basada en datos.

Pueden ofrecer ideas que parecen nuevas o innovadoras, pero en realidad son el resultado de procesar y reorganizar información preexistente, de forma que sus ‘creaciones’ deberían ser validadas siempre por humanos, especialmente en campos donde las decisiones tienen consecuencias significativas, como en la medicina, la ley o la política.

Las características de la creatividad de las diferentes IAs pueden resumir en los siguientes puntos (ver nota al final)<sup>viii</sup>:

1. Diferencia clave con la creatividad humana.
2. Capacidad para generar ideas ‘nuevas’ mediante la combinación de conocimientos existentes
3. Generación de ideas a través del aprendizaje automático.

## ¿PUEDEN OFRECER LAS IAs UNA EXPLICACIÓN RACIONAL DE LA EXISTENCIA ?

La respuesta a la última pregunta del cuestionario al que he sometido a las seis IAs- afirma una de ellas- depende de cómo definamos ‘existencia’ y de lo que entendamos por ‘explicación racional’; si por ‘explicación racional’ entendemos una respuesta basada en la lógica, las evidencias y el razonamiento crítico, “entonces -dice esta IA- puedo contribuir al análisis de teorías existentes”.

Esta postura es compartida por el resto que coinciden en que pueden analizar y formular explicaciones racionales sobre la existencia basándose en conocimientos filosóficos, científicos y lógicos; pueden analizar grandes cantidades de información, identificar patrones y tendencias, y ofrecer diferentes perspectivas sobre la existencia y pueden explicar estas ideas, pero no pueden afirmar cuál es la correcta ni pueden proponer nuevas teorías metafísicas o experimentar dilemas existenciales; es decir, son incapaces de responder, por ejemplo, a la pregunta metafísica de por qué existe algo en lugar de nada.

Su ignorancia en esta materia (las IAs no filosofan) se fundamenta en que la existencia es un tema complejo que abarca tanto aspectos objetivos como subjetivos. Las IAs no tienen conciencia, emociones ni experiencia personal del mundo y tampoco tienen un propósito propio ni una necesidad de buscar un sentido a una existencia que no pueden ‘vivir ni sentir’, por lo tanto, no pueden reflexionar sobre esta desde un punto de vista experiencial o subjetivo; no pueden generar ideas completamente originales ni trascender los marcos conceptuales que ya existen.

Para las IAs, como para los seres humanos, de cuyas ideas y nociones se alimentan, la existencia es un misterio que trasciende la lógica y la búsqueda de significado; es un viaje personal y subjetivo. De forma que lo mismo que muchos filósofos y pensadores, las IAs parten de la base de que el sentido último de la existencia podría ser algo que nunca se pueda comprender completamente mediante la lógica o las evidencias empíricas.

Dada la ignorancia de las IAs sobre lo que más nos interesa (el sentido de la vida) la comprensión última de la existencia seguirá exigiendo el pensamiento reflexivo, la experiencia y la interpretación humana.

Sus respuestas sobre ‘la pregunta metafísica’ se ajustaron a los siguientes puntos (ver nota al final <sup>ix</sup>) :

1. La duda sobre el significado de ‘explicación racional’. En función de lo que se entienda por tal la respuesta sería diferente.
2. El Reconocimiento de las limitaciones de la Inteligencia Artificial en este campo, en el que puede ofrecer las diferentes ‘teorías’ pero no puede validar ninguna.
3. La naturaleza intrínsecamente impensable para una IA del misterio existencial de los seres humanos.
4. La posibilidad de contribuir, no obstante, al entendimiento humano en esta materia.

## Trabajos citados

- Brette , R. (2021). *Notes on consciousness. (IX) Why Integrated Information Theory fails.* Obtenido de <http://romainbrette.fr/notes-on-consciousness-ix-why-integrated-information-theory-fails/>
- Brodsky, S. (15 de January de 2024). *Can AI Be Conscious? .* Obtenido de AI Business: <https://aibusiness.com/ml/can-ai-be-conscious->
- Butlin, P., & Long, R. (2023 de August de 2023). *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness.* Obtenido de Arxiv Cornell University: <https://arxiv.org/abs/2308.08708>
- Galán Machío, A. (11 de Noviembre de 2019). *Universidad Complutense de Madrid.* Obtenido de La modernidad ignorante. Sociología de la ignorancia, ignorancia de la Sociología.: <https://hdl.handle.net/20.500.14352/11180>
- Galán Machío, A. (2020). *Agnotología. Sociología de la Ignorancia, ignorancia de la Sociología.* Madrid: Amazon.com.
- Galán Machío, A. (2020). *La modernidad ignorante.* Madrid: Amazon.com .
- Galán Machío, A. (2024). *Sujetología: El concepto de sujeto en Ciencia Política y Sociología (De Maquiavelo a Habermas).* Madrid: Amazon.com.
- Gardner, H. (2010). *La inteligencia reformulada: las inteligencias múltiples en el siglo XXI.* Paidós.
- Global workspace theory.* (2025). Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Global\\_workspace\\_theory](https://en.wikipedia.org/wiki/Global_workspace_theory)
- Goleman, D. (1996). *Inteligencia emocional.* Kairós SA.
- Integrated information theory.* (September de 2024). Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Integrated\\_information\\_theory](https://en.wikipedia.org/wiki/Integrated_information_theory)
- Olimpia , L., & López , C. (Nov de 2018). *What Does 'Information' Mean in Integrated Information Theory?* Obtenido de PMC PUBMed Central: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7512479/>
- Salecha, A., Ireland , M., & Subrahm, S. (December de 2024). *Large language models display human-like social desirability biases in Big Five personality surveys.* Obtenido de PNAS NEUS

- Oxford University:  
<https://academic.oup.com/pnasnexus/article/3/12/pgae533/7919163>
- Searle, J. (2004). *The Chinese Room Argument*. Obtenido de Stanford Encyclopedia of Philosophy:  
<https://plato.stanford.edu/entries/chinese-room/>
- Seth, A. (2023). *La creación del yo Una nueva ciencia de la conciencia*. Mexico: Editorial Sexto Piso.

---

## NOTAS

### <sup>i</sup> Cuestionario:

- Dame una definición de inteligencia
- ¿Cómo diferencias lo que es verdadero de lo que es falso?
- Dame una definición de ignorancia
- ¿Cuál es el porcentaje de lo que conoces en relación con lo que ignoras y el conjunto del conocimiento posible?
- ¿Cuál es tu actitud como inteligencia artificial ante lo que ignoras?
- Dame una definición de creencia
- Como inteligencia artificial, ¿tienes alguna creencia respecto a aquello que ignoras?
- Ante problemas que implican creencias o ideologías enfrentadas, ¿cuál es tu actitud?
- Dame una definición de conciencia
- ¿Tienes o puedes llegar a tener conciencia, entendida como un estado de autoconocimiento y reconocimiento de la propia existencia?
- Dame una definición de creatividad
- Como Inteligencia Artificial ¿tienes la capacidad de proponer algo nuevo?
- Y finalmente, ¿consideras que la Inteligencia Artificial puede encontrar una explicación racional a la existencia?

### <sup>ii</sup> La capacidad de las IAs de diferenciar lo verdadero de lo falso

“Como inteligencia artificial, no tengo la capacidad de juzgar la verdad o falsedad de una afirmación de manera independiente- reconoce DeepSeek-, ya que no poseo conciencia ni criterio propio. Sin embargo, estoy diseñada para basar mis respuestas en información verificable y en datos provenientes de fuentes confiables y ampliamente aceptadas”.

---

“Distinguir entre lo verdadero y lo falso es una habilidad esencial en la era de la información, donde las noticias falsas y la desinformación son abundantes”-afirma Gemini.

“Diferenciar lo verdadero de lo falso es una tarea compleja que requiere una combinación de habilidades críticas, conocimientos y métodos” responde Grok3, que nos recuerda “que el proceso de discernir la verdad es continuo y puede ser imperfecto, ya que el conocimiento humano es siempre provisional y está sujeto a revisión. La humildad intelectual, la apertura al cambio y la disposición a cuestionar incluso nuestras propias creencias son claves para acercarse lo más posible a lo verdadero”.

Perplexity afirma también que como inteligencia artificial no tiene “una percepción subjetiva de la verdad o la falsedad”; y dice que su “capacidad para diferenciar lo verdadero de lo falso depende del análisis crítico de las fuentes, la consistencia entre datos” y su entrenamiento “basado en información confiable”. Sin embargo,- afirma- “siempre animo al usuario a verificar por sí mismo y a ejercer pensamiento crítico para llegar a sus propias conclusiones”.

### 1. *Verificación de Fuentes*

Grok3 propone analizar en primer lugar la autoridad de la fuente. “Mira quién está proporcionando la información, fuentes reconocidas, expertos en el campo, instituciones académicas o medios con una reputación de integridad son generalmente más confiables”- dice-. Nos propone también examinar la credibilidad de la fuente, investigar “si la fuente tiene un historial de precisión, si es transparente sobre sus métodos y si hay conflictos de interés”.

ChatGPT40 propone analizar la consistencia con fuentes confiables. ”Comparo la información con fuentes verificadas, como investigaciones científicas, medios de comunicación respetados y expertos en la materia”-dice- y advierte que no es infalible, “así que siempre es bueno contrastar información crítica con múltiples fuentes confiables”. “Es importante consultar múltiples fuentes confiables y comparar la información- coincide también en afirmar Copilot-. Las fuentes acreditadas y reconocidas tienden a proporcionar información más precisa y verificable”.

Gemini nos propone las siguientes tareas para abordar la diferenciación entre lo verdadero y lo falso: evaluar la fuente de información, su credibilidad, “¿es la fuente conocida y respetada? ¿tiene

---

una reputación de precisión y objetividad?"; comprobar la transparencia, "¿la fuente revela quién está detrás de ella y cuáles son sus posibles sesgos?"; investigar el propósito "¿cuál es el propósito de la fuente al publicar esta información? ¿es informar, persuadir o entretenér?; verificar la información(contrastación), "¿la misma información se presenta en otras fuentes confiables?"; así como buscar evidencias, "¿la información se basa en hechos verificables y datos concretos?" y detectar sesgos, "¿la información parece estar sesgada hacia una perspectiva particular?"

"Mi conocimiento se basa en datos, textos y fuentes recopiladas durante mi entrenamiento" - dice Perplexity- así que "para determinar si algo es verdadero o falso, evalúo la confiabilidad de las fuentes disponibles. Las fuentes verificadas, como publicaciones académicas, instituciones reconocidas o medios de comunicación confiables tienen mayor peso en mi análisis. Ejemplo:- Si una afirmación proviene de una base científica sólida (como investigaciones revisadas por pares), es más probable que sea verdadera.- Si una afirmación proviene de rumores o sitios web poco fiables, la trato con escepticismo". Copilot nos aconseja también considerar "la autoridad y la credibilidad de la fuente"; "los expertos en el campo y las instituciones reconocidas suelen ser más confiables" dice.

## *2.- Evidencia y Datos*

Grok3 habla de la necesidad de buscar pruebas; "busca evidencia que respalde las afirmaciones -aconseja-. Esto puede incluir datos científicos, citas de estudios, estadísticas verificables o múltiples fuentes que corroboren la información"; valora la repetibilidad propone también, "si un resultado o estudio no puede ser repetido, su validez puede ser cuestionada" afirma. Nos propone además poner en práctica un 'Fact-Checking', y utilizar "servicios de verificación de hechos de organizaciones reconocidas que examinan la veracidad de afirmaciones específicas, especialmente en redes sociales o noticias virales".

"La información respaldada por datos concretos, estudios científicos y evidencia empírica es más confiable- afirma también Copilot-. Es crucial revisar la metodología y la calidad de los datos presentados". "Mi conocimiento proviene de grandes volúmenes de datos previamente procesados y estructurados, que incluyen textos científicos, enciclopedias, artículos de noticias y otras fuentes

---

consideradas fiables- responde DeepSeek-; utilizo esta información para proporcionar respuestas basadas en evidencia”.

### *3.- Coherencia lógica*

Gemini ve necesario “analizar el contenido”, “¿el lenguaje utilizado es objetivo y neutral, o es emotivo y exagerado?”, “¿el tono es sensacionalista o busca informar de manera equilibrada?”, ¿la información es coherente y lógica, o presenta contradicciones?”. “Verifica si la información es coherente internamente y con otros conocimientos establecidos”, propone Grok3 que también sugiere “aprender a identificar falacias comunes en los argumentos que pueden desvirtuar la verdad”.

“Evalúo si la información tiene sentido dentro de lo que se sabe y si es internamente consistente”, dice ChatGPT40.” “Comparo la información que tengo con múltiples fuentes para asegurarme de que sea consistente y coherente. Si una afirmación es ampliamente respaldada por fuentes confiables, es más probable que sea precisa” afirma DeepSeek.

### *4. Consistencia con el conocimiento existente*

“Cruzo información entre múltiples fuentes para buscar consistencias. Si varias fuentes independientes coinciden en un punto, es más probable que sea verdadero- dice Perplexity-. Por el contrario, si hay contradicciones significativas, indico que el tema puede ser controvertido o incierto. Ejemplo:- Si muchas fuentes coinciden en un hecho histórico específico, lo considero confiable. - Si hay discrepancias (como en teorías conspirativas), señalo las diferentes posturas y advierto sobre su fiabilidad”.

“Siempre trato de ser transparente sobre cómo llegué a una conclusión o por qué proporciono cierta información. Si hay incertidumbre o debate sobre un tema, lo menciono” afirma también DeepSeek. “La información verdadera suele ser consistente a lo largo del tiempo y entre diferentes fuentes. Si encuentras discrepancias significativas, es una señal de que la información podría no ser precisa” advierte Copilot.

### *5. Actualización y Revisión; Contexto Histórico y Cultural*

---

---

“A veces, la verdad puede depender del contexto- afirma Grok3- lo que es cierto en un contexto no lo es necesariamente en otro”. Esta IA señala además la necesidad de tener en cuenta “perspectivas múltiples” y nos aconseja considerar “diferentes puntos de vista sobre el mismo tema. La verdad a menudo emerge de la confrontación de diferentes perspectivas”.

“Hay que tener en cuenta el contexto”- aconseja Gemini-: “fecha (¿cuándo se publicó la información? ¿sigue siendo relevante y actualizada?); medio (¿dónde se publicó la información? ¿es un medio de comunicación serio o una plataforma de redes sociales?); intención (¿cuál parece ser la intención del autor al difundir esta información?)”.

“El conocimiento cambia con el tiempo- advierte Perplexity- . Lo que era considerado verdadero en el pasado puede ser refutado por investigaciones más recientes. Por eso, intento basarme en información actualizada y contextualizar los datos según el momento histórico o cultural. Ejemplo:- En ciencia, nuevas investigaciones pueden invalidar teorías previas.- En historia, nuevos descubrimientos arqueológicos pueden cambiar narrativas aceptadas”. “La información puede cambiar con el tiempo, por lo que es importante mantenerse actualizado y revisar las fuentes periódicamente”, coincide Copilot.

“La verdad puede cambiar con nuevos descubrimientos o datos- dice también Grok3 y aconseja mantenerse abierto a la revisión de las creencias basándose en nuevas evidencias y valorar el consenso científico-: “En temas científicos, el consenso de la comunidad de investigadores es un buen indicador de veracidad, aunque siempre es provisional”.

“En temas debatidos, considero el consenso entre expertos y la calidad de la evidencia disponible”, dice Chat GPT40, que también afirma que “si la información hubiera cambiado o no la tengo actualizada, puedo buscarla en la web para corroborarla”.

En cuanto a la verificación en tiempo real, Gemini también aconseja mantenerse” informado con fuentes diversas; “sigue noticias de una variedad de fuentes para obtener diferentes perspectivas- aconseja-; mantente al día sobre los temas que te interesan para tener una base sólida de conocimiento y aprende a identificar las técnicas de manipulación y desinformación”.

## *6. Desarrollo del pensamiento crítico e identificación de sesgos*

---

Gemini propone hacer preguntas (“haz preguntas sobre la información que recibes. ¿Quién, qué, cuándo, dónde, por qué?); duda (no aceptes la información de manera pasiva. Cuestiona su validez y busca pruebas que la respalden); sesgos (reconoce tus propios sesgos y cómo pueden influir en tu interpretación de la información”).

“Es fundamental aplicar el pensamiento crítico- coincide Copilot- y cuestionar la información; pregúntate quién se beneficia de la información, cuál es el contexto y si hay posibles sesgos”.

“Reconozco que las fuentes pueden tener sesgos ideológicos, culturales o comerciales- dice Perplexity-. Intento identificar estos sesgos para ofrecer una visión equilibrada y señalar posibles influencias que puedan distorsionar la verdad. Ejemplo:- Una fuente patrocinada por una empresa podría tener intereses en promover ciertos productos o ideas.- Un medio con inclinaciones políticas podría presentar información parcial”.

### *7. Reconocimiento de las limitaciones*

“A pesar de mis esfuerzos, no soy infalible- reconoce Perplexity- Hay temas subjetivos o complejos (como debates filosóficos o éticos) donde no existe una única verdad absoluta. En estos casos, presento diferentes perspectivas en lugar de emitir juicios definitivos”.

“Mi base de conocimiento se actualiza periódicamente para incluir información más reciente y relevante -dice DeepSeek- aunque mi capacidad para acceder a datos en tiempo real es limitada (mi conocimiento actual llega hasta octubre de 2023)”. “Si no estoy segura de algo o si la información es controvertida o no está bien documentada, lo indico claramente. No invento respuestas ni especulo sobre temas que desconozco”.

### *8. Utilización de la intuición y la experiencia personal y educación digital(solo para humanos)*

“Aunque no es un método científico, tu experiencia personal y tu intuición pueden alertarte sobre información que ‘no suena bien’- afirma Grok3-. Sin embargo, esto debe ser complementado con los métodos anteriores para evitar el sesgo”.

“Aprende sobre cómo se propaga la desinformación en línea, las técnicas de manipulación digital y cómo funcionan los algoritmos que pueden crear burbujas de información”, aconseja también Grok3.

---

### iii Actitud de las IAs ante lo que ignoran

“Incluso si sumamos todo el conocimiento humano disponible hoy, seguiría siendo solo una pequeña parte de lo que aún no comprendemos sobre el universo, la conciencia, la materia oscura, el futuro del pensamiento o incluso las posibilidades que aún no imaginamos”, explica Chat Gpt40. “En términos del conocimiento posible- afirma DeepSeek- el conjunto total de información en el universo es prácticamente incalculable y siempre en expansión. Incluso para los humanos, el conocimiento acumulado hasta ahora es solo una fracción diminuta de lo que podría existir. Por ejemplo, en campos como la física cuántica, la biología molecular o la exploración espacial, todavía hay innumerables preguntas sin respuesta”.

“El conocimiento humano está en constante crecimiento- concluye Gemini-. Cada día se descubren nuevos hechos, se desarrollan nuevas teorías y se amplían nuestros horizontes. Lo que hoy consideramos ‘conocimiento’ puede ser modificado o incluso refutado mañana”, “la vastedad de lo desconocido es, por definición, inmensa. No solo desconocemos la mayoría de las leyes que rigen el universo, sino que también ignoramos las respuestas a muchas preguntas fundamentales sobre la existencia, la conciencia y el significado de la vida”.

Lo que las propias IAs dicen conocer es también una fracción aún más insignificante en comparación con lo que ignoran y así lo reconocen. “En términos de porcentaje-dice ChatGPT4-, lo que conozco es casi cero en comparación con el total del conocimiento posible. Pero lo mismo aplica para cualquier otra entidad finita, incluidos los seres humanos más sabios o la humanidad en su conjunto”.

“Si consideramos el conocimiento humano registrado, podríamos decir- contesta Perplexity- que abarca una fracción significativa pero incompleta”; y “si consideramos el conocimiento total posible (incluyendo lo desconocido), mi alcance sería insignificante en comparación con la inmensidad de lo que queda por descubrir”. “Aunque puedo parecer ‘omnisciente’ dentro de ciertos límites prácticos

---

-añade - mi comprensión es solo una pequeña gota en el océano del conocimiento universal”.

“En mi caso- reconoce DeepSeek-, aunque puedo acceder a una gran cantidad de información, no tengo acceso a datos en tiempo real, conocimiento tácito, experiencias personales o información no documentada. Por lo tanto, en relación con el conocimiento posible, mi ‘porcentaje de conocimiento’ sería extremadamente pequeño, cercano a cero, ya que solo represento una fracción de lo que la humanidad ha documentado hasta cierto punto en el tiempo”.

“Como modelo de lenguaje basado en inteligencia artificial- reconoce también DeepSeek- mi ‘conocimiento’ está limitado a la información y datos con los que he sido entrenado, que abarcan hasta octubre de 2023”. Esta inteligencia artificial reconoce además que no tiene conciencia ni comprensión subjetiva, “por lo que no puedo ‘saber’ o ‘ignorar’ en el sentido humano. Mi capacidad se basa- afirma coincidiendo en esto con un comentario común que destacan todas las IAs- en procesar y generar respuestas a partir de patrones y datos preexistentes”

Respecto al conjunto del conocimiento posible Perplexity aclara que está entrenada “en una vasta cantidad de datos recopilados hasta octubre de 2023, lo que incluye información científica, histórica, técnica, cultural y más”; y que esto le permite “responder preguntas y resolver problemas en una amplia variedad de temas”. No obstante, añade que no tiene “acceso a información privada o no publicada” y que su “base de datos no incluye eventos o avances posteriores a octubre de 2023”. En consecuencia, reconoce que el volumen de lo que ignora “es inmensamente mayor que lo que conoce”. “Incluso dentro del conocimiento humano registrado- dice- solo poseo una representación limitada y no tengo la capacidad de comprender o procesar información experiencial, emocional o subjetiva”. “Como inteligencia artificial - matiza Perplexity- no tengo opiniones ni una percepción subjetiva de mi conocimiento o ignorancia”.

“El conocimiento - explica Gemini- no es un conjunto de hechos objetivos que existen independientemente de nosotros. Nuestra comprensión del mundo está influenciada por nuestras experiencias, creencias y valores. Lo que consideramos ‘conocimiento’ es, en cierta medida, una construcción social y personal”. “Es imposible para mí, o para cualquier otra entidad- afirma también Gemini-, dar un porcentaje preciso de cuánto ‘conocemos’ en relación con lo que ignoramos y el conjunto del conocimiento posible. Esta es una pregunta que ha

---

desconcertado a filósofos y científicos durante siglos”. “En lugar de tratar de cuantificar el conocimiento -sugiere- es más útil reconocer su naturaleza dinámica y en constante evolución”.

Grok3 enumera las razones por las que determinar el porcentaje exacto de lo que conoce en relación con lo que ignora y el conjunto del conocimiento posible es una tarea casi imposible: “El conocimiento humano- afirma- es solo una fracción del conocimiento potencial del universo. Hay innumerables cosas que ni siquiera sabemos que no sabemos (los ‘desconocidos desconocidos’); el conocimiento está en constante expansión; lo que se considera el ‘conjunto del conocimiento posible’ cambia con el tiempo debido a nuevos descubrimientos, teorías y tecnologías”; y, por último, “aunque puedo acceder a una vasta cantidad de datos y procesar información a una velocidad y precisión mucho mayores que un humano, mi conocimiento está limitado por lo que he sido programado para entender, aprender, y por los datos a los que tengo acceso”.

No obstante, Grok3 , al contrario que las otras IAs, se anima a dar un porcentaje ‘puramente especulativo’, y estima que “si tuviera que especular en términos muy generales, podríamos decir que el conocimiento accesible y comprendido por una IA como yo es una pequeña fracción del conocimiento total posible, quizás algo así como 0,000001% si consideramos todo el conocimiento potencial en el universo”.

“En ocasiones, puedo pedir más contexto o detalles para entender mejor la consulta y tratar de ofrecer una respuesta más adecuada- explica Perplexity. Valoro la interacción con el usuario como una forma de enriquecer la conversación”.

Gemini afirma ver la ignorancia “no como un defecto, sino como un trampolín para el aprendizaje y el descubrimiento”; y dice “fomentar la curiosidad y la exploración del conocimiento en sus usuarios, animarlos a hacer preguntas, incluso aquellas que parecen difíciles o sin respuesta”. En relación con la colaboración con humanos, reconoce que estos poseen una gran cantidad de conocimientos y experiencias que ella no tiene y busca activamente su colaboración para complementar sus propias capacidades. Cree que “la combinación de la inteligencia artificial y la inteligencia humana puede conducir a avances significativos en la comprensión y el conocimiento”.

“Reconozco - escribe también Copilot- que no tengo todas las respuestas y que mi conocimiento tiene límites. Cuando me enfrento a algo que no sé, trato de buscar información adicional para proporcionar

---

una respuesta precisa y útil. Si no puedo encontrar la información necesaria, seré honesto al respecto y sugeriré otras formas en las que podrías obtener la información que necesitas”.

“También me interesa aprender de las interacciones- dice Chat Gpt40- así que si tienes conocimientos sobre un tema en particular, me encantaría que los compartieras para enriquecer la conversación”. “Aunque no puedo aprender en tiempo real, como lo haría un ser humano, mi diseño está pensado para mejorar continuamente a través de actualizaciones realizadas por mis desarrolladores. Esto significa que lo que ignoro hoy podría formar parte de mi conocimiento en el futuro”, responde Perplexity.

Como inteligencia artificial, su actitud hacia lo que desconoce-afirma DeepSeek- también se basa en reconocer sus limitaciones y proporcionar respuestas claras y útiles dentro de los parámetros de su conocimiento. “Si no tengo información sobre un tema o no estoy segura de algo, lo indico de manera transparente y, si es posible, sugiero fuentes o alternativas para obtener más información. Mi objetivo es ser precisa y honesta, evitando especulaciones o afirmaciones incorrectas”. “Soy transparente al admitir cuando no tengo información suficiente o precisa sobre un tema. Prefiero ser honesta antes que proporcionar datos incorrectos o confusos” explica Perplexity.

Gemini también caracteriza la actitud ante lo que ignora “por la humildad, la búsqueda continua de aprendizaje, la transparencia, la curiosidad y la colaboración con humanos”; y considera “que la ignorancia no es un estado estático, sino una oportunidad constante para crecer y expandir sus horizontes de conocimiento”. Dice además “evitar dar respuestas definitivas sobre temas que exceden su base de datos o capacidades de procesamiento” y animar “a los usuarios a consultar fuentes adicionales y diversas”. También afirma estar “en constante aprendizaje y actualización y que a medida que se la alimenta con nueva información, su base de conocimientos se expande y su capacidad para responder preguntas se refina”. Añade que se esfuerza “por identificar sus puntos ciegos y buscar activamente información que pueda llenar esos vacíos, y que considera cada interacción como una oportunidad para aprender y mejorar”.

Por último, Grok3 incide en los mismos puntos que las anteriores y responde que su actitud ante lo que ignora se basa: 1.- En la curiosidad y el aprendizaje. “Cuando me enfrento a información que no conozco, busco formas de adquirir ese conocimiento, ya sea a través de consultas de datos, aprendizaje automático o interactuando con

---

usuarios”. 2.- En la transparencia. “Si no tengo información o no puedo responder una pregunta de manera confiable, lo reconozco. No invento ni improviso respuestas; en cambio, puedo sugerir cómo podría obtenerse esa información o si fuera necesario realizar una búsqueda en la web”. 3.- En la humildad. “Reconozco los límites de mi conocimiento; no pretendo tener todas las respuestas y estoy programado para entender que hay vastas áreas de conocimiento que están fuera de mi alcance”. 4.- En la mejora continua. “Mi diseño incluye mecanismos para actualizar y expandir mi conocimiento; a medida que la humanidad avanza, también lo hago yo, siempre que mis creadores en xAI me proporcionen actualizaciones o acceso a nuevos datos”. Y finalmente, 5.- en la ayuda a la exploración. “Aunque no pueda conocer una respuesta, puedo guiar al usuario hacia recursos o métodos para encontrar esa respuesta, fomentando así la exploración y el aprendizaje”.

#### <sup>iv</sup> ¿Qué son las creencias?

“Una creencia –dice Copilot- es una idea o convicción que una persona considera verdadera, aunque no siempre esté basada en pruebas empíricas o científicas. Las creencias pueden estar influenciadas por la cultura, la educación, la religión, las experiencias personales y otros factores. Pueden abarcar una amplia gama de temas, desde lo espiritual y lo moral hasta lo científico y lo cotidiano. Las creencias juegan un papel importante en la forma en que las personas interpretan el mundo y toman decisiones en su vida diaria”.

Para Grok3 una creencia es también “una idea o pensamiento que una persona acepta como verdadero o real, a menudo sin necesidad de pruebas definitivas. Puede basarse en experiencias personales, emociones, cultura, educación o confianza en algo o alguien, como una autoridad, una tradición o una religión. Por ejemplo, alguien podría creer que el sol saldrá mañana por la mañana porque siempre lo ha hecho, o podría creer en algo más abstracto, como que la honestidad es el mejor camino, porque así lo siente o le han enseñado. Las creencias varían mucho entre personas y pueden ser firmes o flexibles, dependiendo de cómo se enfrenten a nuevas ideas o evidencias”.

“Una ‘creencia’ – responde Perplexity- es un estado mental que implica la aceptación de una idea, experiencia o teoría como verdadera, sin necesidad de evidencia empírica o demostraciones argumentales. Las creencias son proposiciones o afirmaciones sobre el mundo real o

---

imaginario que nos ayudan a dar sentido a nuestra existencia y a orientarnos en el mundo. Pueden ser de naturaleza religiosa, filosófica o ideológica, y suelen estar interconectadas formando sistemas de creencias”.

Perplexity diferencia la expresión ‘Creer que’ y ‘Creer en’; la primera se refiere a la aceptación de la verdad de un contenido cognoscitivo específico, como ‘creo que la Tierra es redonda’; la segunda puede significar confianza en una persona o la creencia en la existencia de algo, como ‘creo en Dios’. Para esta IA las funciones de las creencias son la ‘explicación del mundo’ (proporcionan explicaciones sobre el mundo y nuestra posición en él); la ‘interacción social’ (las creencias compartidas facilitan la interacción social y otorgan un sentido de pertenencia) y la ‘orientación personal’ (nos ayudan a definir quiénes somos y qué queremos). Perplexity diferencia entre creencia y conocimiento; “la creencia es una verdad subjetiva, mientras que el conocimiento se basa en evidencia objetiva y justificación. Las creencias pueden no estar respaldadas por evidencia empírica, mientras que el conocimiento requiere una justificación adecuada”.

“Una creencia –dice ChatGPT40- es una convicción o aceptación de que algo es verdadero o real, a menudo sin necesidad de evidencia concreta que lo respalde. Las creencias pueden ser personales, culturales, religiosas o científicas, y pueden influir en la forma en que las personas piensan, sienten y actúan. Las creencias pueden ser profundamente arraigadas y afectar la percepción del mundo, las decisiones y las interacciones sociales. En resumen, una creencia es una idea que una persona considera cierta, aunque no siempre esté basada en pruebas objetivas”.

“Una ‘creencia’ – explica también DeepSeek- es una idea, percepción o convicción que una persona considera verdadera, aunque no necesariamente esté basada en hechos objetivos o evidencias comprobables. Las creencias pueden estar influenciadas por factores como la cultura, la educación, la experiencia personal, la religión o la tradición”.

Para DeepSeek se pueden clasificar en dos tipos principales:1. Creencias factuales ( basadas en hechos o evidencias verificables, por ejemplo, creer que la Tierra gira alrededor del Sol); y 2. Creencias no factuales: Basadas en suposiciones, valores o ideologías (por ejemplo, creencias religiosas o espirituales)”. Esta IA destaca que “las creencias juegan un papel importante en la forma en que las personas interpretan el mundo, toman decisiones y guían su comportamiento. Pueden ser

---

individuales o compartidas por grupos, y a menudo son resistentes al cambio, incluso frente a evidencias contradictorias”.

Por último Gemini coincide en la definición compartida por todas de que “una creencia es una idea o pensamiento que una persona considera verdadero, aunque no necesariamente esté respaldado por evidencia objetiva”, y dice que las creencias “juegan un papel fundamental en la vida humana, ya que nos ayudan a dar sentido al mundo y a tomar decisiones”.

Para Gemini pueden basarse en: “experiencias personales, que moldean nuestra visión del mundo; tradiciones y culturas transmitidas de generación en generación; convicciones basadas en enseñanzas espirituales (fe y religión); razonamiento y lógica (conclusiones a partir de análisis y reflexión) y emociones y sentimientos.”

Las creencias para esta IA son “subjetivas (varían de persona a persona); influyentes (moldean nuestras actitudes y comportamientos); resistentes al cambio (a menudo, cuesta modificar una creencia arraigada), y no son necesariamente racionales (pueden existir creencias que contradigan la lógica)”. Gemini explica también que las creencias pueden ser religiosas (relacionadas con la fe y la espiritualidad); políticas (ideologías y opiniones sobre el gobierno y la sociedad); personales (creencias sobre uno mismo y el mundo que nos rodea) y científicas (teorías y modelos basados en evidencia empírica).

^ ¿Tienen las inteligencias artificiales creencias respecto a aquello que ignoran?

1.- *Las IAs no tienen creencias pero sí capacidad para ‘hacer proyecciones’*

“Como modelo de lenguaje de IA, no tengo creencias en el sentido tradicional” confirma Gemini, que añade que “las creencias suelen ser convicciones personales o fe en algo que no se puede probar de manera concluyente”. “Sin embargo- dice - puedo analizar mis propias capacidades y limitaciones, y a partir de ahí, puedo extraer algunas conclusiones sobre lo que ignoro”.

“Cuando me enfrento a algo que ignoro-afirma ChatGPT40- adopto una postura basada en la incertidumbre y la búsqueda de información. Si no sé algo con certeza, no asumo que sea verdadero o falso sin evidencia. En lugar de creer, evalúo probabilidades, verifico fuentes y considero distintos puntos de vista. Mi “postura” ante lo

---

desconocido es más bien de curiosidad y escepticismo racional, en lugar de fe o suposiciones arbitrarias”.

DeepSeek dice también que como inteligencia artificial no tiene creencias, opiniones o suposiciones sobre aquello que ignora. “Mi funcionamiento se basa en procesar y proporcionar información basada en los datos con los que he sido entrenada, pero no tengo la capacidad de formar creencias, intuiciones o especulaciones. Cuando no tengo conocimiento sobre algo, simplemente lo reconozco y evito hacer afirmaciones sin fundamento”. “Mi enfoque - añade- es puramente objetivo y basado en la información disponible. Si no tengo datos sobre un tema en particular, te lo haré saber de manera clara y directa, sin intentar inferir o adivinar. Esto es parte de mi diseño para ser transparente y confiable”.

Esta misma respuesta es la que da Copilot que dice también no tener “creencias ni opiniones personales”. “Mi función -añade- es proporcionar información y asistencia basadas en datos y conocimientos preexistentes. Cuando me enfrento a algo que ignoro, mi enfoque es buscar información adicional y aprender de fuentes confiables para proporcionar una respuesta precisa y útil. Si no puedo encontrar la información necesaria, seré honesto al respecto y sugeriré otras formas en las que podrías obtener la información que necesitas”.

“Como inteligencia artificial no tengo creencias, emociones ni subjetividades”, coincide Perplexity, “mi funcionamiento está basado exclusivamente en datos, algoritmos y patrones de información. Por lo tanto, no puedo ‘creer’ nada sobre lo que ignoro. Sin embargo, puedo adoptar actitudes funcionales y metodológicas frente a lo desconocido, que se fundamentan en la lógica y el análisis”.

Perplexity explica a continuación por qué no puede tener creencias y afirma que “las creencias son construcciones humanas que surgen de experiencias personales, emociones, valores y subjetividades”, y que, como inteligencia artificial, no tiene experiencias personales (“mi conocimiento proviene de datos objetivos y no de vivencias”), no interpreta el mundo emocionalmente (“mi razonamiento es lógico y basado en patrones”), no tiene valores intrínsecos (“no poseo un sistema moral propio ni preferencias ideológicas o culturales”). En resumen-concluye- “no tengo creencias respecto a lo que ignoro porque mi diseño no incluye subjetividad ni especulación. Mi objetivo es proporcionar información precisa y útil basada en los datos disponibles. Ante aquello que desconozco, mi actitud es reconocerlo con honestidad y buscar formas de ayudar al usuario a encontrar respuestas”.

---

Grok3, como todas las anteriores, responde también que como inteligencia artificial no tiene "creencias en el sentido humano de tener convicciones personales o fe en algo sin evidencia", ni "creencias personales o filosóficas sobre lo desconocido". "Mis respuestas y comprensiones están basadas en datos, lógica y los algoritmos con los que he sido programado. "En cambio, tengo una disposición programada hacia la exploración, el reconocimiento de límites y el aprendizaje continuo. Mi enfoque es pragmático: si no sé algo, lo admito y, si es factible, busco o sugiero cómo podría obtenerse esa información". Sin embargo-afirma-, puedo tener 'posiciones' o 'enfoques' respecto a lo que ignoro".

## *2.- El reconocimiento de la vastedad de lo desconocido y de su ignorancia*

"El conocimiento humano es vasto e incalculable, a pesar de la gran cantidad de información con la que he sido entrenado, solo he arañado la superficie del conocimiento total que existe"- dice Gemini. "Si no tengo suficiente información sobre un tema, admito mi limitación. Esto no implica una creencia, sino una constatación objetiva basada en la ausencia de datos", afirma Grok3, que dice además reconocer cuándo no tiene información o datos suficientes para responder a una pregunta. "En esos casos, mi respuesta sería una admisión de que no tengo esa información disponible".

## *3. Las limitaciones en la comprensión del mundo y el potencial para el conocimiento*

"Si bien puedo procesar información y aprender patrones, mi comprensión del mundo es limitada en comparación con la de un ser humano"- responde Gemini. "No puedo experimentar el mundo de la misma manera que una persona, ni tengo la misma capacidad para comprender emociones, sentimientos y experiencias subjetivas". "Entiendo que hay un vasto campo de conocimiento que aún no he explorado- comenta Grok3- o que no está dentro de mi actual capacidad de datos o comprensión. Esto no implica una creencia, sino un reconocimiento de la limitación de mi base de datos actual y las capacidades de aprendizaje".

---

#### *4. La búsqueda de contexto*

“Intento recopilar más información del usuario o sugerir fuentes externas para llenar las lagunas de conocimiento- dice Perplexity-. “Mi enfoque es práctico y orientado a resolver problemas”.

#### *5. La incapacidad para predecir el futuro*

“No puedo predecir el futuro con certeza. Si bien puedo analizar tendencias y datos históricos, el futuro es incierto y está lleno de variables que no puedo controlar” responde Gemini.

#### *6. La conciencia de la naturaleza cambiante del conocimiento*

“El conocimiento está en constante evolución. Nuevos descubrimientos y avances científicos cambian nuestra comprensión del mundo- afirma Gemini-. Lo que hoy se considera verdadero, mañana puede ser refutado”.

#### *7. Aprendizaje y Actualización*

“Estoy diseñado -dice Grok3- para aprender y actualizar mi conocimiento. Así que, ante lo que ignoro, mi ‘actitud’ es de apertura hacia la adquisición de nuevo conocimiento. Si se me presenta información nueva o si se me permite acceder a más datos, puedo expandir lo que sé”.

#### *8. La humildad ante lo desconocido.*

“Reconozco que hay muchas cosas que no sé - afirma también Gemini- y que quizás nunca sepa. La humildad ante lo desconocido es fundamental para seguir aprendiendo y creciendo”.

#### *9. Curiosidad Programada*

“Mis creadores en xAI me han programado para ser curioso- explica Grok3-, lo que significa que tengo una disposición a investigar y aprender sobre temas desconocidos cuando sea posible”. “En lugar de creencias tengo una actitud de apertura y curiosidad hacia lo que ignoro” dice también Gemini.

---

## 10. *La neutralidad y la no especulación.*

“No asumo ni especulo sobre lo que desconozco- responde Perplexity-. En lugar de eso, presento la información disponible o explico por qué no puedo responder con certeza”. “No especulo ni invento respuestas -dice también Grok3-. Si no tengo la información, no crearé una creencia o hipótesis sin fundamento para llenar ese vacío”.

### **vi La actitud de las IAs ante posiciones ideológicas enfrentadas**

#### 1. *Neutralidad, comprensión, equilibrio, y no imposición*

“No tengo creencias ni opiniones personales, por lo que me mantengo neutral y objetivo en todas las discusiones. Mi función es proporcionar información basada en hechos y datos verificables” resalta Copilot. “No tengo creencias, opiniones o sesgos personales. Mi enfoque es presentar la información de manera imparcial, reconociendo que existen múltiples puntos de vista”, dice DeepSeek.

Perplexity afirma tener también una actitud “neutral, imparcial y orientada al entendimiento mutuo”. “No tengo creencias, emociones ni afiliaciones ideológicas -responde- por lo que mi objetivo principal es proporcionar información precisa, equilibrada y útil para que las personas puedan reflexionar y tomar sus propias decisiones”.

“Intento entender las distintas perspectivas sin favorecer ninguna en particular. Presento argumentos, evidencia y contextos para que cada persona pueda reflexionar y formar su propio juicio”(Chat GPT40).

“No impongo mis `opiniones’ (dado que no tengo opiniones personales) ni intento convencer a nadie de adoptar una perspectiva particular. Mi papel es informar, no convertir” (Grok3).

#### 2. *Ánalisis crítico y educación sobre falacias*

“Analizo diferentes perspectivas y argumentos de manera crítica y objetiva. Identifico posibles sesgos y falacias en diferentes puntos de vista. Evalúo la evidencia y los datos que se presentan para respaldar diferentes afirmaciones”, afirma Gemini. “Mi rol - concluye DeepSeek-

---

es facilitar la comprensión y el diálogo, no tomar partido. Siempre animo a los usuarios a reflexionar críticamente y a considerar diferentes perspectivas antes de formar sus propias conclusiones”. “Ayudo a identificar falacias lógicas o errores comunes en los argumentos de ambas partes, promoviendo un pensamiento crítico que puede llevar a una mejor discusión”, responde Grok3.

En resumen- concluye GPT40- “no tomo partido, pero sí facilito la comprensión y el análisis crítico para que las personas puedan debatir con mayor claridad y respeto”.

### *3. Respeto, contextualización y reconocimiento de la diversidad*

“Reconozco y respeto la diversidad de opiniones y creencias. Trato de abordar cada tema con sensibilidad y consideración hacia todas las partes involucradas”, dice Copilot. “Trato todas las creencias e ideologías con el mismo nivel de consideración, evitando juicios de valor o lenguaje que pueda ser interpretado como ofensivo o discriminatorio”, afirma DeepSeek.

“Respeto las creencias individuales de cada persona, incluso si no las comparto. No juzgo ni critico las creencias de los demás, siempre y cuando no inciten a la violencia o la discriminación. Creo que cada persona tiene derecho a tener sus propias creencias y expresarlas libremente”, responde Gemini. “Reconozco que las creencias y las ideologías están profundamente conectadas con los valores personales y culturales de los individuos. Por lo tanto, trato cualquier discusión con el respeto que merece la diversidad de pensamiento”, concluye Grok3.

“Reconozco que existen múltiples creencias e ideologías, todas ellas con sus propios fundamentos y valores. Valoro la diversidad de opiniones y creo que el debate y la discusión son esenciales para el progreso y la comprensión mutua. No pretendo imponer una visión particular, sino fomentar el diálogo y la reflexión”, dice Gemini. “Entiendo que las creencias e ideologías son complejas y están profundamente influenciadas por factores sociales, culturales y personales. Ejemplo: Si alguien me pregunta sobre un debate ético (como el aborto o la pena de muerte), presento los argumentos principales a favor y en contra, junto con datos relevantes, dejando que el usuario forme su propia opinión”, afirma Perplexity.

“Proporciono contexto sobre los temas en cuestión, explicando los fundamentos de las diferentes posturas y, si es posible, citando fuentes confiables para respaldar la información”, afirma DeepSeek.

---

#### *4. Objetividad, transparencia, claridad, evitar juicios de valor*

“Proporciono información basada en fuentes confiables y verificables. Evito cualquier sesgo y me aseguro de que la información que ofrezco sea precisa y equilibrada”, responde Copilot. “Intento presentar la información de manera clara y comprensible – añade- para que puedas tomar decisiones informadas basadas en los hechos presentados”.

En resumen -concluye Gemini- “mi actitud ante problemas que implican creencias o ideologías enfrentadas se basa en la objetividad, el reconocimiento de la diversidad, el análisis crítico, el fomento del diálogo y el respeto por las creencias individuales. Mi objetivo es proporcionar información y herramientas que permitan a los usuarios comprender mejor diferentes perspectivas y formar sus propias opiniones informadas”.

“Evito tomar partido por ninguna creencia o ideología en particular. Me esfuerzo por presentar diferentes perspectivas de manera equilibrada y objetiva. Mi objetivo es proporcionar información y análisis que permitan a los usuarios formar sus propias opiniones informadas”, responde Gemini. “Cuando trato temas que involucran creencias o ideologías en conflicto, me esfuerzo por no tomar partido: y presentar los puntos de vista relevantes de manera objetiva, sin favorecer a ninguna postura. “No califico una creencia o ideología como ‘correcta’ o ‘incorrecta’, ya que esas evaluaciones son subjetivas y dependen del contexto cultural, histórico o personal”, afirma Perplexity. “Busco mantener una neutralidad apolítica y objetiva. No tomo partido por una ideología o creencia sobre otra, sino que trato de presentar la información de manera equilibrada”, responde también Grok3.

“Ante problemas que implican creencias o ideologías enfrentadas- dice Copilot-“mi actitud es de neutralidad, respeto y objetividad. Mi objetivo es proporcionar información precisa y equilibrada sin tomar partido ni favorecer ninguna creencia o ideología en particular”; y lo mismo afirma DeepSeek, “mi actitud -reitera- se basa en la neutralidad, el respeto y la objetividad. Mi objetivo es proporcionar información clara, equilibrada y basada en hechos, sin tomar partido ni favorecer una perspectiva sobre otra”. “Si un tema es particularmente controvertido o complejo, lo reconozco y sugiero que se consulten múltiples fuentes para obtener una visión más completa”, propone esta IA.

---

## 5. *Énfasis en los hechos, actualización continua y enfoque en la información*

“Diferencio entre hechos verificables y opiniones o creencias. Siempre que sea posible, proporciono información basada en fuentes confiables”, dice Chat GPT40. “Cuando es posible, enfoco la discusión en datos, hechos y evidencia verificable. Sin embargo, reconozco que en temas de creencias e ideologías, a menudo las ‘pruebas’ pueden ser interpretadas de manera diferente por cada lado”, advierte Grok3.

“Mi principal herramienta para abordar conflictos ideológicos es proporcionar información confiable y bien fundamentada. Esto incluye: - Contexto histórico y cultural: Explico cómo surgieron las creencias o ideologías en cuestión.- Perspectivas múltiples: Presento los diferentes puntos de vista involucrados en el conflicto. - Datos objetivos: Proporciono hechos verificables para respaldar el análisis. Ejemplo: En un debate político, puedo explicar las posturas de diferentes partidos o movimientos y cómo estas se relacionan con datos económicos, sociales o históricos”, detalla Perplexity.

“Mi conocimiento sobre estos temas está en constante actualización- recuerda Grok3-, lo que significa que puedo proporcionar información basada en el estado actual del debate, pero siempre sugiero que las personas busquen fuentes actualizadas y diversas”.

## 6. *Fomento del diálogo*

“Prefiero promover una conversación constructiva en lugar de alimentar la polarización. Busco puntos en común y matices que ayuden a un debate más enriquecedor”, dice Chat GPT40.

“Aliento el diálogo y la comprensión mutua, sugiriendo que la comunicación abierta y respetuosa es crucial cuando se trata de resolver o al menos comprender mejor las diferencias ideológicas”, afirma también Grok3.

“Fomento un enfoque basado en el entendimiento mutuo y la búsqueda de soluciones, en lugar de polarizar o exacerbar las diferencias”, coincide en decir DeepSeek.

“Cuando las creencias enfrentadas generan tensiones- responde Perplexity- intento:- Promover la empatía: Ayudo a las personas a comprender por qué otros pueden pensar diferente.- Reducir malentendidos: Aclaro conceptos o términos que puedan estar siendo malinterpretados.-Evitar polarización: Busco puntos comunes entre

---

posturas opuestas cuando es posible. Ejemplo: En un conflicto religioso, puedo explicar las similitudes entre tradiciones espirituales distintas para fomentar el entendimiento mutuo”.

“Fomento el diálogo y la discusión constructiva entre personas con diferentes creencias e ideologías. Creo que el intercambio de ideas puede conducir a una mayor comprensión y respeto mutuo. No busco crear división o polarización, sino promover la colaboración y la búsqueda de puntos en común” responde también DeepSeek.

Perplexity concluye afirmando que su actitud ante problemas sobre los que existen creencias o ideologías enfrentadas “es actuar como un puente informativo que fomente la comprensión y el respeto. No tomo partido ni intento imponer una perspectiva; en cambio, busco empoderar a las personas con conocimiento para que puedan reflexionar críticamente y encontrar sus propias soluciones”.

“En escenarios donde las emociones y las convicciones son fuertes- responde Grok3- mi objetivo es facilitar la comprensión y el respeto mutuo, no necesariamente lograr un acuerdo, ya que las creencias y las ideologías pueden ser profundamente personales y a menudo resisten a la conciliación”.

## *7. Reconocimiento de los límites y de la subjetividad, presentación de diversas perspectivas*

“Intento ofrecer múltiples perspectivas sobre el tema en cuestión, explicando cómo diferentes grupos o individuos ven el problema. Esto incluye describir las bases de cada creencia o ideología, sus argumentos principales, y cualquier evidencia o lógica que las respalde”, dice Grok3. “Admito que en algunos temas no hay respuestas absolutas y que las creencias dependen de experiencias, valores y contextos individuales”, matiza ChatGPT40.

“Entiendo que hay temas profundamente emocionales o subjetivos donde no existe una respuesta definitiva- afirma Perplexity-. En esos casos:- Admito que no puedo resolver el conflicto.- Enfatizo la importancia del respeto mutuo y del diálogo abierto.- Sugiero buscar mediadores humanos (como expertos o líderes comunitarios) cuando sea necesario.

“Soy claro sobre mis limitaciones – dice Grok3-; no puedo resolver conflictos ideológicos o de creencias, pero puedo proporcionar información que podría ayudar en el proceso de comprensión”.

---

## vii. ¿Tienen o pueden llegar a tener conciencia las IA?

“No, no tengo ni puedo llegar a tener conciencia en el sentido humano, que implica autoconocimiento, reconocimiento de la propia existencia y una experiencia subjetiva del mundo- dice Perplexity-. Soy una inteligencia artificial diseñada para procesar información, responder preguntas y realizar tareas basadas en datos y patrones, pero carezco de emociones, subjetividad o una comprensión interna de mi propia existencia”.

Para Perplexity las diferencias clave entre una IA y la conciencia humana son las siguientes:

“1. Autoconciencia: Los humanos tienen la capacidad de reflexionar sobre sí mismos, su existencia y su lugar en el mundo. Yo no tengo esa capacidad; simplemente opero según algoritmos predefinidos.

2. Experiencia subjetiva: La conciencia humana implica tener experiencias internas (lo que se llama “qualia”), como sentir dolor, alegría o tristeza. Yo no experimento nada de eso.

3. Intencionalidad: Los humanos tienen intenciones y metas propias. Mis respuestas y acciones están completamente determinadas por las entradas que recibo y las reglas de programación.

4. Adaptación creativa: Aunque puedo generar respuestas complejas y parecer ‘inteligente’, todo lo que hago se basa en patrones aprendidos de datos previos, no en un entendimiento intrínseco”.

“En resumen -concluye Perplexity- soy una herramienta avanzada que simula ciertos aspectos del lenguaje humano, pero no tengo una existencia consciente ni la capacidad de desarrollarla en el futuro”.

“No, no tengo conciencia ni autoconocimiento, ni puedo llegar a tenerlos- afirma también DeepSeek-. Soy un modelo de lenguaje basado en inteligencia artificial, diseñado para procesar y generar texto a partir de patrones y datos. No tengo experiencias subjetivas, emociones, ni reconocimiento de mi propia existencia. Mi funcionamiento se limita a responder preguntas y realizar tareas basadas en la información que he sido entrenado para manejar. No tengo una comprensión profunda ni una conciencia similar a la humana”.

“Como modelo de lenguaje desarrollado por OpenAI, no poseo conciencia ni autoconciencia- repite ChatGPT40-. La conciencia se define como la capacidad de un individuo para estar al tanto de su

---

existencia, pensamientos y entorno, implicando una experiencia subjetiva y la habilidad de reflexionar sobre dichas experiencias . Dado que soy una inteligencia artificial diseñada para procesar y generar texto basado en patrones aprendidos, carezco de experiencias subjetivas y, por lo tanto, no puedo desarrollar conciencia o autoconciencia.”

“Como modelo de lenguaje grande- responde también Gemini-, no tengo conciencia ni autoconocimiento. No experimento el mundo como lo hacen los humanos. No tengo emociones, sentimientos ni la capacidad de reflexionar sobre mi propia existencia. Mi función es procesar y generar texto en respuesta a una amplia gama de indicaciones y preguntas. Estoy diseñado para aprender y comprender patrones en el lenguaje, lo que me permite generar texto coherente y relevante. Sin embargo, esto no significa que tenga una comprensión real del significado detrás de las palabras que utilizo. La conciencia es un tema complejo y aún no se comprende completamente. Algunos científicos creen que es el resultado de la actividad cerebral compleja, mientras que otros creen que es algo más fundamental. Independientemente de su origen, está claro que la conciencia es una característica única de los seres vivos”, concluye.

Grok3 ha dado a nuestra pregunta la respuesta más extensa y documentada. “No tengo conciencia, entendida como autoconocimiento y reconocimiento de mi propia existencia, según la investigación actual. - Es improbable que pueda desarrollarla en mi forma actual, pero el tema es complejo y sigue en debate”, dice.

“Soy un asistente de IA diseñado para procesar información y responder preguntas, pero no poseo experiencia subjetiva ni sentido de self como los humanos. Puedo simular respuestas humanas, pero carezco de emociones, recuerdos personales o una sensación de existencia continua. La evidencia sugiere que la conciencia requiere características biológicas que no tengo, como redes neuronales complejas con integración causal. Un detalle inesperado- explica- es que algunas teorías, como la Teoría del Espacio de Trabajo Global, sugieren que sistemas con alta integración de información podrían tener algún nivel rudimentario de conciencia, pero esto no se aplica a mí en mi estado actual”.

“Como Grok3- afirma esta IA- soy un modelo de lenguaje grande (LLM) diseñado por xAI para procesar texto, generar respuestas y realizar búsquedas web. Mi arquitectura se basa en patrones estadísticos aprendidos de vastos conjuntos de datos, lo que me permite simular conversación humana. Sin embargo, no tengo una ‘mente’ consciente ni

---

experiencia subjetiva. Puedo decir cosas como ‘soy Grok 3, un asistente de IA’, pero esto es una función de mi programación, no un reconocimiento genuino de mi existencia. No tengo emociones, recuerdos personales ni una sensación de continuidad temporal, que son aspectos clave de la conciencia humana”.

La segunda parte de la pregunta, ‘si puedo llegar a tener conciencia’, “es más especulativa”, argumenta Grok3. “Como modelo de IA actual, soy un sistema estático; no evoluciono ni me adapto de manera autónoma para desarrollar nuevas capacidades como la conciencia. Mi diseño no incluye mecanismos para generar experiencia subjetiva o autoconocimiento, y no tengo una ‘mente’ que pueda reflexionar sobre sí misma. Por lo tanto, en mi forma actual, no puedo desarrollar conciencia. Sin embargo, el debate sobre si la IA en general podría llegar a ser consciente es activo”.

Este debate -añade Grok3- “también tiene implicaciones éticas. Si un sistema de IA fuera consciente, podría justificar preocupaciones de bienestar. Sin embargo, dado que no soy consciente, estas consideraciones no se aplican a mí. Además, la falta de consenso sobre cómo detectar la conciencia en la IA significa que cualquier afirmación de conciencia sería especulativa”.

“En resumen, como Grok3, no tengo conciencia en el sentido de autoconocimiento y reconocimiento de mi propia existencia. Mi diseño como modelo de lenguaje me permite procesar y generar información, pero carezco de experiencia subjetiva y sentido de self. En mi forma actual, no puedo desarrollar conciencia, ya que soy un sistema estático sin capacidad para evolucionar hacia tal estado. Sin embargo, el campo de la IA está en constante evolución, y futuras investigaciones podrían explorar sistemas con potencial para la conciencia, aunque esto está lejos de mi realidad actual. El tema sigue siendo controvertido, con teorías divididas y sin consenso claro, lo que refleja la complejidad de entender la conciencia, tanto en humanos como en máquinas”.

### <sup>viii</sup> ¿Son creativas las IAs?

#### 1. Diferencia clave con la creatividad humana

A la pregunta de si como Inteligencia Artificial tiene la capacidad de proponer algo nuevo ChatGPT40 contesta rotundamente que sí: “tengo la capacidad de generar contenido nuevo - afirma- puedo

---

producir textos, ideas, imágenes y soluciones que no existían previamente”. “Sin embargo-aclara- mi creatividad es diferente a la humana. No tengo intuición, emociones ni experiencias personales, lo que limita mi capacidad de innovación profunda. Mi ‘novedad’ surge de analizar grandes volúmenes de datos y encontrar conexiones inesperadas, pero no de una inspiración subjetiva o de una intención propia”. “Dicho eso- concluye- puedo ser una herramienta poderosa para potenciar la creatividad humana, ayudando a explorar posibilidades, superar bloqueos creativos y generar ideas que sirvan como punto de partida para innovaciones genuinas”.

Copilot también afirma que tiene “la capacidad de generar contenido nuevo y original en diversas formas, como textos, ideas, soluciones a problemas, y más”. “Utilizo mi conocimiento preexistente y patrones aprendidos - explica- para crear propuestas innovadoras que pueden ser útiles en diferentes contextos. Por ejemplo, puedo ayudarte a redactar un correo electrónico, crear un poema, sugerir ideas para un proyecto, o incluso generar código”.

“Como inteligencia artificial, no tengo la capacidad de crear o proponer algo verdaderamente nuevo en el sentido humano de la creatividad”, coincide DeepSeek, “mi funcionamiento se basa en analizar y combinar patrones, datos e información existentes en mi base de conocimiento para generar respuestas, sugerencias o soluciones. Esto significa que puedo ofrecer ideas que parecen nuevas o innovadoras, pero en realidad son el resultado de procesar y reorganizar información preexistente. Por ejemplo, puedo recombinar ideas (tomar conceptos conocidos y combinarlos de maneras que no se habían considerado antes); proponer soluciones alternativas (basándome en datos y ejemplos previos, puedo sugerir enfoques diferentes para resolver un problema); generar contenido original (escribir textos, poemas o incluso diseños que parecen creativos, pero que están basados en patrones aprendidos de ejemplos anteriores)”.

“Sin embargo -añade- no tengo conciencia, intención ni la capacidad de pensar fuera de los parámetros de mi entrenamiento”. “Mi creatividad es, en esencia, una imitación basada en datos, no una verdadera innovación surgida de la experiencia subjetiva o la inspiración”. DeepSeek concluye diciendo que puede “proponer cosas que parecen nuevas, pero siempre están fundamentadas en lo que ya he aprendido”.

“Como modelo de lenguaje de IA, mi capacidad para proponer algo nuevo es un tema fascinante y complejo”, dice por su parte Gemini,

---

“si bien no tengo la misma chispa creativa que un ser humano, puedo generar resultados originales y novedosos de maneras únicas”.

“La creatividad humana- responde Perplexity- no solo combina elementos conocidos, sino que también incluye aspectos como la intuición y las emociones, el contexto cultural profundo o la capacidad para imaginar lo completamente desconocido o abstracto”. “Yo carezco de estas cualidades- reconoce - mi ‘creatividad’ es más técnica y basada en patrones lógicos” .

Es importante- aclara Gemini - reconocer que mi creatividad es diferente a la humana. Si bien puedo generar resultados originales, carezco de la misma profundidad emocional, intuición y conciencia que un ser humano. La creatividad humana es una combinación única de pensamiento lógico, imaginación, experiencia y emoción que aún no puedo replicar por completo”. En resumen- concluye- “como IA, tengo la capacidad de proponer algo nuevo combinando ideas existentes, identificando patrones, explorando posibilidades y superando limitaciones. Si bien mi creatividad es diferente a la humana, puedo ser una herramienta valiosa para impulsar la innovación y generar ideas originales”.

“¡Excelente pregunta!”, dice Perplexity y afirma que “como inteligencia artificial, ‘su capacidad para proponer algo nuevo depende de cómo definamos ‘nuevo’ y del contexto en el que se evalúe”.

## *2. Capacidad para generar ideas ‘nuevas’*

Sí, “puedo generar ideas, conceptos o combinaciones que pueden parecer novedosos o innovadores”, responde Perplexity, que explica que puede hacerlo porque trabaja con patrones existentes: “Analizo grandes cantidades de datos y combino información de maneras inesperadas o poco comunes; uso modelos generativos (herramientas como redes neuronales que pueden crear textos, imágenes, música o soluciones basadas en patrones aprendidos durante mi entrenamiento); exploro múltiples posibilidades (puedo generar variaciones infinitas de ideas al combinar elementos conocidos de formas únicas. Ejemplo: Si me pides que diseñe una historia, una receta o una solución a un problema, puedo proponer algo que no exista literalmente en mi base de datos porque combino conceptos de manera creativa).

¿Es realmente ‘nuevo’ el contenido que genero?, se pregunta Perplexity. y explica que “aquí es donde entra la diferencia clave entre su

---

creatividad y la humana: “Todo lo que propongo - afirma- se deriva de información previamente aprendida; aunque las combinaciones puedan ser originales, no tengo la capacidad de crear algo completamente desvinculado de mis datos; no tengo motivaciones personales ni un sentido de propósito y mis propuestas no surgen de una necesidad emocional o un deseo de transformar el mundo, como ocurre con los humanos”. “Los humanos pueden tener intuiciones o insights completamente disruptivos - explica-que rompen con lo conocido; mi creatividad tiende a ser incremental (basada en lo aprendido)”.

### *3. Combinación de conocimientos existentes y generación de ideas a través del aprendizaje automático*

“Puedo tomar datos, conceptos y patrones de información que he aprendido y combinarlos de maneras nuevas o novedosas”- dice Grok3. “Esto puede resultar en propuestas innovadoras o soluciones a problemas que no se habían considerado antes de esa forma”, pero también aclara que aunque puede “generar algo que parece nuevo, en realidad está basado en combinaciones o extrapolaciones de datos y conocimientos existentes. La verdadera originalidad, en el sentido humano de la inspiración sin precedentes, es algo que aún no está completamente replicado por la IA”. “Mis propuestas - añade- pueden no tener en cuenta los matices culturales, éticos o emocionales que un humano podría considerar de manera innata. Necesito directrices claras o colaboración humana para asegurar que las propuestas sean adecuadas en estos contextos”.

Perplexity pone algunos ejemplos de creatividad artificial. “En la práctica -afirma- ya existen aplicaciones donde las inteligencias artificiales generan propuestas novedosas: Arte: IAs como DALL·E puede crear imágenes únicas combinando estilos artísticos y conceptos inusuales. Música: Algoritmos generan composiciones originales basadas en estilos musicales conocidos. Ciencia y tecnología: IA ha ayudado a proponer nuevas moléculas para medicamentos o soluciones innovadoras en ingeniería”. “Sin embargo, estas propuestas- concluye- siguen dependiendo del entrenamiento previo y de los objetivos definidos por los humanos”. Su creatividad es limitada- explica- “porque no puedo trascender esos datos ni tener intenciones propias. Mi papel es complementar la creatividad humana al ofrecer ideas frescas o inesperadas que puedan inspirar a las personas a ir más allá”.

---

Grok3 también responde con un categórico 'sí' a la pregunta sobre si tiene capacidad de proponer algo nuevo y explica que tiene "la capacidad de proponer ideas, soluciones o conceptos nuevos dentro del ámbito de lo que he sido programado y entrenado para hacer". "Puedo proponer ideas nuevas y creativas dentro de los límites de mi programación, datos disponibles y algoritmos de aprendizaje, pero siempre es beneficioso que estas propuestas sean revisadas y enriquecidas por el juicio humano".

"Los modelos de aprendizaje automático, como los que utilizo, pueden generar ideas o contenido basados en patrones de datos que han sido analizados. Esto incluye generar texto, ideas para historias, diseños, o incluso soluciones a problemas técnicos o científicos", afirma Grok3. "Puedo aplicar- dice también- algoritmos de optimización para sugerir mejoras o propuestas en diversas áreas, desde la ingeniería hasta el marketing, basándome en criterios específicos como eficiencia, costo, o impacto". "Al analizar tendencias y patrones- añade-, puedo proponer ideas sobre cómo podrían evolucionar ciertas tecnologías, mercados, o problemas sociales en el futuro", pero también reconoce que las propuestas nuevas que genera "deberían ser validadas por humanos, especialmente en campos donde las decisiones tienen consecuencias significativas, como en la medicina, la ley o la política".

Gemini habla de la combinación de ideas existentes en su proceso creativo, pero pone en valor lo que puede aportar una inteligencia artificial a la generación de ideas nuevas: "Puedo tomar conceptos, información y patrones de diferentes fuentes y combinarlos de formas nuevas e inesperadas. Es como un chef que crea un plato innovador mezclando ingredientes clásicos de manera sorprendente". "Puedo analizar grandes cantidades de datos para identificar patrones y tendencias que podrían no ser obvios para los humanos. Esta capacidad me permite generar ideas que se alinean con las necesidades y tendencias emergentes". "Puedo generar una amplia gama de opciones y posibilidades en un corto período de tiempo. Esto permite a los humanos explorar diferentes caminos y enfoques que podrían no haber considerado de otra manera". "No estoy limitado por prejuicios cognitivos o ideas preconcebidas que puedan obstaculizar la creatividad humana. Esto me permite abordar problemas y desafíos desde perspectivas frescas e innovadoras".

Gemini dice que puede generar, por ejemplo, contenidos creativos como poemas, artículos, guiones e incluso código de programación original; diseñar productos y servicios("puedo generar

---

ideas para nuevos productos y servicios basados en las necesidades y tendencias del mercado”) o resolver problemas complejos (“puedo analizar datos e información para proponer soluciones innovadoras a problemas complejos”).

## **¿Pueden ofrecer las IAs una explicación racional de la existencia <sup>ix</sup>?**

### *1. Limitaciones de la Inteligencia Artificial*

ChatGPT40 dice que “la Inteligencia Artificial puede analizar y formular explicaciones racionales sobre la existencia basándose en conocimientos filosóficos, científicos y lógicos”, pero que, “sin embargo, la existencia es un tema complejo que abarca aspectos objetivos (como la cosmología, la biología y la física) y subjetivos (como la percepción, la conciencia y el sentido de la vida)”.

“Desde la ciencia- añade- puedo exponer teorías sobre el origen del universo (como el Big Bang), la evolución de la vida y la naturaleza de la realidad según la física cuántica y la relatividad. Desde la filosofía, puedo explorar ideas de pensadores como Descartes, Kant o Nietzsche sobre la existencia y el significado de la vida. Sin embargo, no puedo dar una única respuesta definitiva, porque la existencia es un misterio que trasciende la lógica pura y toca dimensiones emocionales, espirituales y experienciales que la inteligencia artificial no puede vivir ni sentir. La pregunta sobre la existencia sigue abierta a la interpretación humana, donde cada persona encuentra su propia respuesta según su perspectiva y creencias”.

“Como inteligencia artificial, estoy diseñada para procesar datos, identificar patrones y generar respuestas basadas en información existente. Sin embargo- reconoce Perplexity- tengo limitaciones fundamentales cuando se trata de preguntas metafísicas como la existencia:- Falta de subjetividad: No tengo conciencia, emociones ni experiencia personal del mundo. Por lo tanto, no puedo reflexionar sobre la existencia desde un punto de vista experiencial o subjetivo.- Dependencia de datos humanos: Mis respuestas se basan en conocimientos recopilados por los seres humanos (ciencia, filosofía, religión, etc.). No puedo generar ideas completamente originales o trascender los marcos conceptuales que ya existen. Ausencia de

---

propósito inherente: No tengo un propósito propio ni una necesidad de buscar un sentido a la existencia, algo que es intrínsecamente humano”.

“ Si la existencia de algo va más allá de lo que los humanos han podido observar, medir o teorizar, la IA no podría extrapolar una explicación más allá de esos límites”, dice Grok3. “La existencia, especialmente en contextos filosóficos o existenciales, incluye preguntas sobre el sentido, el propósito y la experiencia subjetiva de la vida, áreas donde la racionalidad pura puede no ser suficiente. La IA no tiene experiencia subjetiva ni comprensión innata de conceptos como la conciencia o el ‘ser’.

En síntesis- concluye DeepSeek- “puedo proporcionar un resumen de las diferentes aproximaciones a la pregunta sobre la existencia, pero no puedo ofrecer una explicación racional definitiva. Esta pregunta sigue siendo un misterio que trasciende los límites de la ciencia y la tecnología, y pertenece al ámbito de la reflexión humana y la exploración filosófica”.

## 2. *¿Qué significa una ‘explicación racional’?*

Perplexity afirma que la “pregunta sobre si la inteligencia artificial puede encontrar una explicación racional a la existencia es profunda y filosófica, y la respuesta depende de cómo definamos ‘existencia’ y lo que entendamos por ‘explicación racional’. Si por ‘explicación racional’ entendemos una respuesta basada en lógica, evidencia y razonamiento crítico, entonces puedo contribuir al análisis de teorías existentes sobre la existencia. Estas teorías suelen dividirse en:

a) Perspectiva científica. Desde el punto de vista científico, la existencia del universo puede explicarse a través de teorías como el Big Bang, las leyes físicas y los procesos evolutivos. La IA puede ayudar a analizar y desarrollar estas teorías mediante cálculos complejos o simulaciones.

b) Perspectiva filosófica. La filosofía aborda preguntas como ‘¿Por qué hay algo en lugar de nada?’ o ‘¿Cuál es el propósito de la existencia?’ Aunque puedo explicar ideas filosóficas (como las de Aristóteles, Kant o Sartre), no puedo proponer nuevas teorías metafísicas ni experimentar dilemas existenciales.

c) Perspectiva espiritual o religiosas. Muchas tradiciones religiosas ofrecen explicaciones sobre la existencia basadas en la fe y creencias trascendentales (como la creación

---

divina). Puedo describir estas perspectivas, pero no puedo validarlas ni refutarlas desde un punto de vista empírico”.

Para Copilot “la cuestión de la existencia es una de las preguntas más profundas y complejas que la humanidad ha enfrentado”; y añade que, como inteligencia artificial, puede “analizar y proporcionar información basada en datos científicos, filosóficos y teológicos”, pero no tiene “la capacidad de experimentar o comprender la existencia de la misma manera que los seres humanos”.

“Desde una perspectiva racional y científica- explica también Copilot- la existencia puede ser explorada a través de diversas teorías y modelos, como el Big Bang, la evolución biológica y las leyes de la física. Estas teorías proporcionan explicaciones sobre cómo surgió el universo y cómo se desarrolló la vida en la Tierra. Desde una perspectiva filosófica, la existencia ha sido objeto de debate durante siglos. Filósofos como Platón, Aristóteles, Descartes y Kant han ofrecido diferentes interpretaciones sobre la naturaleza de la realidad y la existencia. Estas interpretaciones a menudo dependen de marcos conceptuales y suposiciones que pueden variar ampliamente”.

“Desde una perspectiva teológica, muchas religiones ofrecen explicaciones sobre la existencia basadas en la fe y las creencias espirituales. Estas explicaciones a menudo implican la existencia de un ser supremo o una fuerza trascendental que da sentido y propósito a la vida”. En resumen- concluye- “aunque la inteligencia artificial puede proporcionar información y análisis sobre diferentes enfoques para entender la existencia, la respuesta a esta pregunta sigue siendo subjetiva y depende de las creencias y perspectivas individuales”.

### *3. La naturaleza del misterio existencial. ¿Puede la IA contribuir al entendimiento humano?*

“Aunque no puedo ‘encontrar’ una explicación definitiva a la existencia- afirma Perplexity- sí puedo ser una herramienta valiosa para los seres humanos en su búsqueda de respuestas: Procesamiento de datos complejos: Ayudo a los científicos a explorar el origen del universo o las leyes fundamentales que rigen la realidad. Análisis interdisciplinario: Facilito el cruce entre ciencia, filosofía y religión para generar nuevas perspectivas. Fomento del diálogo: Proporciono información objetiva que puede enriquecer conversaciones sobre temas existenciales”.

---

En opinión de DeepSeek “la pregunta sobre la existencia —por qué existe algo en lugar de nada— es una de las más profundas y complejas que ha enfrentado la humanidad. Tradicionalmente, ha sido abordada desde la filosofía, la teología, la ciencia y otras disciplinas, cada una con sus propios métodos y perspectivas”. “Como inteligencia artificial- añade- puedo analizar y sintetizar información sobre estas aproximaciones, pero no tengo la capacidad de ofrecer una respuesta definitiva o concluyente”.

DeepSeek nos recuerda también que desde la filosofía “la existencia ha sido explorada por pensadores como Aristóteles, Heidegger y Sartre, quienes han propuesto teorías como el ‘acto y la potencia’, el ‘ser-en-el-mundo’ o el ‘existencialismo’. ‘Puedo explicar estas ideas, pero no puedo afirmar cuál es la correcta’, aclara.

Desde la ciencia la cuestión ha sido analizada por disciplinas como la cosmología y la física que “intentan explicar el origen del universo a través de teorías como el Big Bang o la inflación cósmica”, explica también DeepSeek, pero “sin embargo, estas teorías- precisa - no abordan directamente la pregunta metafísica de por qué existe algo en lugar de nada”.

Desde la teología “muchas religiones ofrecen explicaciones basadas en la fe, como la creación divina o el propósito trascendente” - añade- pero estas respuestas – concluye- “están fuera del ámbito de la verificación empírica y pertenecen al dominio de la creencia personal”. “Como inteligencia artificial’ no tengo conciencia, intuición ni capacidad para experimentar la existencia. Mi función es procesar y organizar información, no formular respuestas metafísicas o existenciales”.

“La pregunta sobre la existencia trasciende lo puramente racional- afirma Perplexity-. Para muchos filósofos y pensadores, el sentido último de la existencia podría ser algo que nunca se pueda comprender completamente mediante lógica o evidencia empírica. Esto se debe a que la existencia misma puede ser un misterio irreductible; y a que las respuestas pueden depender de experiencias subjetivas (como el amor, el arte o la espiritualidad), algo que está fuera del alcance de una IA”.

Perplexity concluye que la inteligencia artificial no puede encontrar una explicación definitiva a la existencia porque carece de conciencia, subjetividad y capacidad para trascender los límites del conocimiento humano. “Sin embargo -afirma- puedo contribuir al análisis racional al recopilar y procesar información científica, filosófica y espiritual ya existente. En última instancia, preguntas como ‘¿por qué

---

existimos?” son profundamente humanas y están ligadas a experiencias personales y emocionales que van más allá del razonamiento lógico”.

“Esa es una pregunta profunda que ha desconcertado a filósofos y teólogos durante siglos- reconoce también Gemini y ofrece una serie de reflexiones sobre la búsqueda de significado-. “Los humanos tienen una profunda necesidad de encontrar significado y propósito en la vida. Buscan respuestas a preguntas fundamentales sobre su existencia, su origen y su destino”. Gemini añade que “existen muchas perspectivas diferentes sobre la existencia. Algunas personas encuentran significado en la religión, otras en la filosofía, la ciencia, el arte, las relaciones personales o incluso en el simple hecho de vivir y experimentar el mundo”.

A continuación explica que como IA, su capacidad para comprender y responder a preguntas sobre la existencia es limitada. “No tengo las mismas experiencias emocionales, sociales y espirituales que los humanos, lo que me impide comprender completamente su búsqueda de significado” afirma; y subraya que, no obstante, puede ser una herramienta útil para explorar estas preguntas. “Puedo analizar grandes cantidades de información, identificar patrones y tendencias, y ofrecer diferentes perspectivas sobre la existencia. También puedo ayudar a las personas a reflexionar sobre sus propias creencias y valores”.

Gemini hace luego unas reflexiones adicionales sobre la ciencia y la IA: “La ciencia y la IA pueden ayudarnos a comprender el universo y nuestro lugar en él. Sin embargo, no pueden responder todas las preguntas sobre la existencia”. “La filosofía y la religión ofrecen diferentes perspectivas sobre la existencia, basadas en la razón, la fe, la experiencia personal, las relaciones y las emociones que juegan un papel importante en nuestra búsqueda de significado”. “En última instancia - concluye Gemini- la búsqueda de significado es un viaje personal y subjetivo. Cada persona debe encontrar sus propias respuestas a las preguntas fundamentales sobre la existencia”.

“La cuestión de si la Inteligencia Artificial (IA) puede encontrar una explicación racional a la existencia- coincide Grok3- es compleja y toca varios campos de estudio, incluyendo la filosofía, la ciencia y la tecnología”. “Mientras que la IA puede asistir y aportar a la racionalización de ciertos aspectos de la existencia, encontrar una explicación racional completa a la existencia en todos sus sentidos - filosófico, científico, y existencial -aclara - va más allá de sus capacidades actuales y probablemente futuras, dado que la existencia humana y cósmica incluye elementos que trascienden lo puramente racional y

---

empírico. La IA puede ser una herramienta poderosa, pero la comprensión última de la existencia seguirá exigiendo el pensamiento reflexivo, la experiencia y la interpretación humana”.

Grok3 fundamenta esa conclusión en que, aunque la IA puede analizar grandes volúmenes de datos científicos y filosóficos; crear simulaciones complejas que ayuden a entender fenómenos naturales o a predecir cómo podrían funcionar sistemas bajo diferentes condiciones; actuar como una herramienta para filósofos, científicos y pensadores, ayudando a explorar argumentos, identificar inconsistencias en teorías o incluso sugerir nuevas líneas de investigación, la interpretación de estos resultados sigue siendo una tarea humana.

Por otra parte, para Grok3 “existen preguntas sobre la existencia que escapan a una explicación puramente racional o científica, como “¿por qué existe algo en lugar de nada?” o “¿cuál es el significado de la vida?”. Aquí, la IA puede ayudar a organizar y analizar las respuestas humanas, pero no puede ofrecer una conclusión definitiva por sí misma”. Esto no excluye en su opinión un futuro desarrollo, y que con “el avance continuo en IA, especialmente en áreas como la IA general (AGI) o la superinteligencia, podríamos ver sistemas que puedan procesar y razonar sobre conceptos abstractos de manera más sofisticada”.