

# ¿Puede una máquina ser consciente? Una respuesta popperiana al problema de las otras mentes en la era de la inteligencia artificial

*Can a machine be conscious? A popperian response to the problem of other minds in the age of artificial intelligence*

ALEJANDRA VALDES-MILLAN\*

Docente Universidad de Caldas, Manizales, Caldas, Colombia  
alejandra.valdes@ucaldas.edu.co

RECIBIDO EL 15 DE JULIO DE 2025, ACEPTADO EL 15 DE OCTUBRE DE 2025

## RESUMEN

El desarrollo de inteligencias artificiales sofisticadas ha reavivado el problema filosófico de las otras mentes: ¿cómo justificamos la atribución de consciencia a entidades no biológicas? ¿Es posible que una máquina sea consciente? Desde una perspectiva popperiana, la consciencia es un fenómeno emergente biológicamente anclado que requiere memoria con historia evolutiva, control plástico autónomo, experiencia subjetiva (Mundo 2) y creatividad argumentativa genuina. Las inteligencias artificiales actuales, en tanto productos del Mundo 3 creados por seres humanos, carecen de estas condiciones constitutivas y, por tanto, no pueden considerarse conscientes. Se articula el análisis del lenguaje ordinario de John L. Austin sobre el problema de las otras mentes con la ontología pluralista de Karl Popper (teoría de los tres mundos) y su teoría evolucionista de la consciencia como fenómeno emergente. Este marco se aplica críticamente a los grandes modelos de

lenguaje contemporáneos, dialogando con la literatura reciente sobre consciencia artificial. Las IA actuales —y previsiblemente cualquier sistema carente de base biológica e historia evolutiva— no cumplen los criterios popperianos para la consciencia genuina. Simulan procesos cognitivos mediante recombinación estadística, pero no poseen experiencia subjetiva ni creatividad argumentativa. La visión popperiana ofrece así una advertencia metodológica: no debe confundirse la simulación de consciencia con su posesión.

## PALABRAS CLAVE

consciencia, emergentismo biológico, inteligencia artificial, Karl Popper, problema de las otras mentes

## ABSTRACT

The development of sophisticated artificial intelligence has reignited the philosophical problem of other minds: how do we

\*  <https://orcid.org/0009-0005-5850-7503> Google Scholar



justify attributing consciousness to non-biological entities? Is it possible for a machine to be conscious? From a Popperian perspective, consciousness is a biologically anchored emergent phenomenon that requires memory with an evolutionary history, autonomous plastic control, subjective experience (World 2), and genuine argumentative creativity. Current artificial intelligences, as World 3 products created by human beings, lack these constitutive conditions and, therefore, cannot be considered conscious. This paper integrates John L. Austin's analysis of ordinary language regarding the problem of other minds with Karl Popper's pluralist ontology (theory of the three worlds) and his evolutionary theory of consciousness as an emergent phenomenon. This framework is

critically applied to major contemporary language models, engaging with recent literature on artificial consciousness. Current AIs – and predictably any system lacking a biological basis and evolutionary history – do not meet Popper's criteria for genuine consciousness. They simulate cognitive processes through statistical recombination, but they possess neither subjective experience nor argumentative creativity. The Popperian view thus offers a methodological warning: the simulation of consciousness must not be confused with its possession.

#### KEY WORDS

consciousness, biological emergentism, artificial intelligence, Karl Popper, problem of other minds

## Introducción

Atribuir estados mentales a otros seres humanos es una práctica cotidiana que realizamos, en la mayoría de los casos, sin mayor cuestionamiento. Esta capacidad nos permite explicar, e incluso predecir, el comportamiento de quienes nos rodean. Sin embargo, cuando llevamos esta atribución más allá del ámbito humano – por ejemplo, hacia animales no humanos o, más recientemente, hacia sistemas de inteligencia artificial (IA) – surgen interrogantes filosóficos profundos: ¿cómo justificamos epistemológicamente la existencia de otras mentes? ¿Es posible que entidades no biológicas, como las IA, posean consciencia? ¿O acaso la consciencia es un fenómeno irreductiblemente ligado a la evolución biológica?

El presente artículo aborda estas preguntas desde la perspectiva de dos filósofos del siglo XX: Karl Popper y John L. Austin. La elección de este marco no es arbitraria. Por un lado, Austin (1989) nos ofrece una herramienta metodológica fundamental al situar el problema de las otras mentes en el análisis del lenguaje ordinario, lejos de especulaciones metafísicas infértiles. Por otro lado, Popper (1997) proporciona una ontología pluralista – los tres mundos: físico (M1), mental (M2) y de los productos objetivos del pensamiento (M3) – junto con una teoría

evolucionista de la consciencia como fenómeno emergente. Esta síntesis permite sostener, como se argumentará, que la consciencia genuina requiere una base biológica y una historia evolutiva, condiciones que las IA actuales no cumplen.

El problema de las otras mentes adquirió particular relevancia a mediados del siglo XX, en el contexto del desarrollo de soluciones a problemas epistemológicos fundamentales. Preguntas como ¿cómo sabemos que otros individuos tienen experiencias conscientes como las nuestras? o ¿podemos realmente conocer los estados mentales ajenos? fueron abordadas por Austin desde el análisis del uso cotidiano del lenguaje. Para Austin, lejos de ser un enigma metafísico irresoluble, esta cuestión surge en nuestra vida diaria a través de expresiones como “siento dolor” o “estoy consciente”. El análisis de estas expresiones — con sus ambigüedades y sutilezas — no nos lleva a descartar la pregunta, sino a comprender mejor la naturaleza de nuestra atribución de consciencia a otros. Austin sostiene que nuestras creencias sobre las mentes ajenas se basan en inferencias razonables a partir de evidencia conductual, verbal y de sentido común, entendido este último como una herramienta para la vida social más que como una fuente infalible de conocimiento.

Popper coincide con Austin en rechazar un escepticismo radical respecto de las otras mentes, pero su contribución va más allá al anclar la consciencia en un proceso evolutivo. Para Popper, la consciencia no es una sustancia separada — como en el dualismo cartesiano — ni un mero epifenómeno — como en el materialismo reduccionista —. Es, más bien, un fenómeno emergente que surge de la interacción entre el mundo físico (M1) y el mundo de los estados mentales subjetivos (M2), y que alcanza su nivel más elevado en los seres humanos gracias a la capacidad de articular pensamientos en forma de enunciados y argumentos, los cuales pasan a formar parte del mundo objetivo del conocimiento (M3). Esta visión pluralista (que incluye interacción entre mundos), que Popper prefiere denominar simplemente “pluralista” (Popper, 1997, p.163), rechaza tanto el paralelismo psicofísico como el epifenomenalismo, pues ambos reducen o ignoran el poder causal de la mente.

Ahora bien, si la consciencia es un producto de la evolución biológica, ¿qué estatus podemos atribuir a las inteligencias artificiales? En las últimas dos décadas, los avances en IA — especialmente los modelos de lenguaje de gran escala (LLMs) y las redes neuronales profundas — han reavivado el debate sobre la posibilidad de consciencia artificial.

Algunos autores, desde posiciones funcionalistas o panpsiquistas, han sugerido que ciertos sistemas computacionales podrían, en principio, ser conscientes (Chalmers, 2023; Bostrom, 2014). Sin embargo, desde el marco popperiano que aquí se defiende, esta posibilidad resulta insostenible: las IA, por más sofisticadas que sean en términos de procesamiento de información, carecen de la historia evolutiva, de la base biológica y de la capacidad de generar creativamente contenidos originales para el Mundo 3. Son, en rigor, productos del Mundo 3 – creaciones humanas –, no habitantes del Mundo 2.

El artículo se organiza de la siguiente manera. En la segunda sección, se expone el problema de las otras mentes tal como lo formula Austin, destacando su método de análisis del lenguaje ordinario como punto de partida para una aproximación no escéptica. En la tercera, se presenta la ontología de los tres mundos de Popper y su defensa del interaccionismo pluralista frente al epifenomenalismo y el paralelismo. En la cuarta, se desarrolla la teoría evolucionista de la consciencia como fenómeno emergente, incluyendo los conceptos de jerarquía de controles y control plástico. En la quinta sección – la más relevante para el debate contemporáneo – se aplican estos criterios a las IA actuales, dialogando con la literatura reciente sobre consciencia artificial. En la sexta, se discuten las implicaciones de esta postura en relación con el funcionalismo computacional y el enactivismo, señalando también las limitaciones del enfoque popperiano. Finalmente, en las conclusiones se sintetizan los hallazgos y se responde explícitamente a la pregunta central. Se sostiene que, desde el marco popperiano, las IA actuales no cumplen las condiciones necesarias para la consciencia genuina y que no debe confundirse la simulación de estados conscientes con su posesión real.

### **El problema de las otras mentes: Austin y el lenguaje ordinario**

El problema de las otras mentes – cómo podemos justificar la existencia de consciencia en otros individuos – adquirió particular relevancia a mediados del siglo XX, en el contexto del desarrollo de soluciones a problemas epistemológicos fundamentales. Austin (1989), uno de sus representantes más destacados, abordó esta cuestión en su ensayo *Otras mentes*, ofreciendo una perspectiva que se distancia tanto del escepticismo radical como de las respuestas metafísicas infértiles.

Para Austin (1989), la pregunta sobre cómo sabemos si otros individuos tienen experiencias conscientes como las nuestras no es un enigma metafísico abstracto. Surge, más bien, en nuestra vida cotidiana a través del uso del lenguaje ordinario. Al analizar expresiones como “siento dolor” o “estoy consciente”, nos encontramos con ambigüedades y sutilezas que dificultan una respuesta simple. Sin embargo, esto no implica que debemos descartar la pregunta como irresoluble. Por el contrario, Austin sostiene que el análisis del lenguaje ordinario es un punto de partida fundamental para comprender el pensamiento y la naturaleza humana, pues nos permite tener una comprensión cercana y amplia de la forma como nos comprendemos a nosotros mismos y podemos comprender a otros.

Austin no se limita al lenguaje. También incorpora la información que obtenemos por los sentidos como parte de la base para nuestras creencias sobre otras mentes. Aunque nuestros sentidos y nuestra capacidad de razonar pueden fallarnos en ocasiones, esto no significa que sean siempre engañosos. Podemos basar nuestras creencias sobre la consciencia de los demás en una amplia gama de evidencias: su comportamiento, sus expresiones verbales y nuestro propio sentido común. Este último, para Austin, es una herramienta para la vida social, no una fuente de conocimiento infalible, pero sí una guía práctica razonable.

La respuesta que Austin ofrece a la pregunta “¿cómo sabemos que otros tienen consciencia?” no es definitiva en términos de una demostración lógica incontrovertible. Pero ello no la invalida. Austin nos invita a reconocer que nuestras creencias sobre la consciencia de los demás se basan en inferencias razonables a partir de la evidencia disponible. No necesitamos una certeza cartesiana para convivir con otras mentes; nos basta con justificaciones prácticas y contextuales que funcionan en nuestra vida diaria.

Popper coincide con esta propuesta en un punto crucial: ser individuos con capacidades racionales no sugiere una distancia radical entre mis ideas y la posibilidad de asumir ideas en otros. Para Popper, al igual que para Austin, la atribución de consciencia a otros seres humanos es una inferencia legítima y necesaria para la vida en común. Sin embargo, como veremos en el siguiente apartado, Popper va más allá al proporcionar una base ontológica y evolutiva para esta atribución: la consciencia no es solo algo que inferimos por conducta y lenguaje, sino un producto

de la evolución biológica que compartimos con otros organismos de nuestra especie.

### **El marco popperiano: tres mundos, interaccionismo y crítica al reduccionismo**

Para comprender la visión popperiana de la consciencia es necesario situarla en el debate clásico sobre la relación entre mente y cuerpo. Popper acepta la existencia tanto de eventos físicos como de estados o eventos mentales, sin caer en ninguna forma de reduccionismo o eliminativismo. Como él mismo señala:

En síntesis, el fisicalismo niega lo evidente, esto es, la existencia de estados mentales o de la consciencia. Entre las teorías que no niegan la existencia evidente de los estados mentales, varias teorías de la mente han rivalizado por conseguir ser aceptadas. (Popper, 1997, p.102)

El dualismo cartesiano postula dos sustancias fundamentales: la mental (*res cogitans*) y la física (*res extensa*). Para Descartes, la consciencia —nuestra capacidad de pensar— es la base de nuestra existencia, llegando a la certeza del *Cogito, ergo sum*. Sin embargo, esta distinción introduce un problema complejo: ¿cómo interactúan dos sustancias radicalmente diferentes? Descartes propuso la glándula pineal como lugar de interacción, pero la dificultad de explicar esta interacción ha sido una de las principales críticas al dualismo cartesiano.

Popper asume el interaccionismo, pero se distancia del dualismo de sustancias. Para él, postular dos sustancias distintas implicaría retroceder a un esencialismo que considera inadmisibile. En lugar de ello, Popper propone una visión que emerge de su teoría de los tres mundos, en la que se asume la existencia tanto de eventos mentales como de eventos físicos sin comprometerse con atributos radicalmente diferentes entre sí.

Una alternativa al interaccionismo es el paralelismo psicofísico, según el cual los procesos mentales y físicos constituyen dos series de eventos completamente independientes pero que guardan una correspondencia perfecta y preestablecida. Los cambios en el mundo físico no causan cambios en la mente; simplemente funcionan de manera paralela.

Popper rechaza esta postura por las siguientes razones. Si adoptamos el paralelismo, tendríamos que asumir que incluso entidades físicas muy básicas, como los electrones, poseen consciencia. Pero, como argumenta Popper:

Tal vez un electrón esté investido de consciencia, pero según la teoría cuántica, todos los electrones son exactamente iguales cualquiera que haya sido su historia. Con otras palabras, no les afecta su historia. Si se asume el paralelismo, significa que, incluso si disponen de consciencia, no pueden disponer de ninguna clase de memoria. Ahora bien, la consciencia sin memoria dispondría de momentos fugaces de consciencia, cada uno de los cuales duraría un periodo mínimo de tiempo, y cada uno de los cuales carecería completamente de conexión con todos los demás. (Popper, 1997, p.162)

Para Popper, la memoria es fundamental para la consciencia. Una consciencia sin memoria – como la que tendría un electrón según el paralelismo – sería una mera sucesión de «chispas» conscientes sin conexión entre sí, lo que no se ajusta a nuestra comprensión intuitiva de lo que significa ser consciente.

El epifenomenalismo es la segunda opción que Popper considera. Según esta perspectiva, los eventos mentales (pensamientos, sentimientos, consciencia) son meros subproductos o «efectos secundarios» de los procesos físicos del cerebro. Los eventos mentales no causan nada en el mundo físico; son un epifenómeno pasivo. Una forma de entenderlo es pensar en una máquina de vapor cuyo silbido acompaña su funcionamiento, pero no influye en él. De manera similar, para el epifenomenalista, la consciencia sería un «silbido» que acompaña a la actividad cerebral sin influir en ella.

Popper rechaza el epifenomenalismo porque no puede reconciliar adecuadamente la experiencia consciente ni explicar la interacción entre mente y cuerpo que considera fundamental para entender la naturaleza humana y el avance de la ciencia. Como veremos a continuación, la existencia de productos objetivos del pensamiento – como las teorías científicas – que influyen de vuelta en el mundo físico demuestra, para Popper, que la consciencia no puede ser un epifenómeno insignificante.

Para superar estas dificultades, Popper propone su célebre teoría de los tres mundos, presentada en su ponencia *Epistemology without a Knowing Subject* (1967; recogida posteriormente en Popper, 1992). Esta ontología pluralista distingue tres ámbitos de la realidad:

el Mundo 1 (M1) comprende el mundo físico: átomos, cuerpos, cerebros y energía. El Mundo 2 (M2) es el mundo de los estados mentales y la experiencia subjetiva, es decir, los pensamientos, sentimientos, percepciones y el dolor. El Mundo 3 (M3) está constituido por los contenidos objetivos del pensamiento con existencia independiente de cualquier mente individual: las teorías científicas, las obras de arte, las matemáticas y los lenguajes

La clave de esta propuesta es que existe una relación dinámica entre los tres mundos. El Mundo 3 —las ideas— puede influir en el Mundo 1 —el mundo físico— a través de la tecnología y la ingeniería. A su vez, nuestras experiencias subjetivas (Mundo 2) pueden dar lugar a nuevas ideas que reposan en el Mundo 3. En palabras de Popper:

El hecho de la existencia de los problemas objetivos que existen en el mundo 3, tales como los problemas que plantea la teoría de los números primos, junto con el hecho de que el descubrimiento de un problema de esa clase puede llevar a grandes y evidentes cambios en el mundo 1, creo que demuestra que el mundo 1 no está cerrado ni completo, sino abierto hacia el mundo 3, en donde el mundo 2 actúa como intermediario. Si esto es así, entonces el epifenomenalismo debe estar equivocado: la mente o la consciencia no puede ser un epifenómeno insignificante. (Popper, 1997, p. 163)

Popper no se autodenomina simplemente «interaccionista», sino más bien pluralista. Debe comprenderse el interaccionismo en su obra como la relación que se da entre los tres mundos, donde a su vez emergen productos nuevos. Existe una interacción entre los diferentes niveles de realidad, y los eventos físicos y mentales interactúan de manera dinámica.

Para Popper, esta interacción es fundamental no solo para comprender la consciencia, sino también para explicar el avance de la ciencia. La consciencia —como parte del Mundo 2— emerge de la interacción con el Mundo 1 a través de la evolución biológica, y sus productos —teorías, desarrollos científicos— se ubican en el Mundo 3, demostrando que la

consciencia no es un mero subproducto pasivo, sino una instancia activa en la generación de conocimiento objetivo.

### **Consciencia como fenómeno emergente y biológico**

Para Popper, la consciencia no es un añadido misterioso ni una sustancia separada. Es, ante todo, un producto de la evolución biológica. A medida que los organismos se volvieron más complejos y desarrollaron cerebros más evolucionados, también desarrollaron capacidades cognitivas más sofisticadas, incluyendo la consciencia. Desde una perspectiva evolutiva, la consciencia habría surgido como una ventaja adaptativa que permite a los organismos interpretar su entorno y adaptarse a él. Popper introduce la noción de jerarquía de controles para explicar cómo la consciencia emerge y funciona en el organismo. Todos los organismos poseen una jerarquía que se divide entre controles básicos y controles superiores:

- Controles básicos: regulan funciones vitales como el latido del corazón, la respiración y los procesos metabólicos. Son automáticos y no requieren un alto nivel de consciencia.
- Controles superiores: se encargan de funciones más complejas como la coordinación de movimientos, la percepción y la toma de decisiones. Son más flexibles y están estrechamente relacionados con la consciencia.

Dentro de los controles superiores, Popper introduce el concepto de control plástico. Este se refiere a aquellos controles que permiten una amplia gama de respuestas posibles ante una misma situación. Los controles plásticos son más flexibles y adaptativos que los controles no plásticos, que ofrecen una respuesta única y predecible. Gracias al control plástico podemos —partiendo de los datos que tenemos— aprender, modificar, retroalimentarnos y ser flexibles con los procesos aprendidos previamente.

La consciencia, según Popper, está estrechamente ligada a estos controles plásticos. La capacidad de elegir entre diferentes opciones, de aprender de la experiencia y de adaptarse a nuevas situaciones son características típicas de los sistemas de control plástico y están íntimamente relacionadas con nuestra experiencia consciente.

La consciencia, en tanto emerge de un proceso evolutivo, no es un fenómeno de todo o nada. Popper reconoce la existencia de grados de consciencia en otros seres vivos. Esto indica que, como seres humanos, nos distinguimos de otros seres vivos no por tener consciencia o no tenerla, sino por el nivel que alcanzamos.

¿Qué nos diferencia? Fundamentalmente, dos capacidades:

1. La autoconsciencia: la capacidad de comprendernos a nosotros mismos, de tener una idea sobre nosotros mismos, el Yo.
2. El lenguaje en sus niveles superiores (descriptivo y argumentativo), que permite comprender nuestra capacidad como individuos de imaginar, crear y producir explicaciones originales y novedosas.

Estas capacidades nos permiten, a partir de nuestra comprensión del Mundo 1 o del Mundo 2, establecer contenidos o resultados que pasan a formar parte del Mundo 3.

Para Popper, lo que consideramos el estado más elevado de consciencia se alcanza cuando pensamos de manera consciente y articulada, es decir, cuando expresamos nuestros pensamientos en forma de enunciados y argumentos. No basta con tener pensamientos; es necesario que estos pensamientos sean claros, ordenados y que podamos expresarlos. Este resultado debe ser creado y originado por nosotros mismos en nuestro nivel de realidad del Mundo 2, no ser el mero resultado de la conjunción de múltiples datos. La imaginación y la creatividad tienen un papel fundamental en este proceso.

Al articular nuestros pensamientos, no solo pensamos, sino que también somos conscientes de que lo estamos haciendo. Adquirimos, así, consciencia de nuestra propia consciencia. En palabras de Popper:

El yo o el ego es el resultado de haber logrado vernos a nosotros mismos desde el exterior y, por tanto, de habernos situado a nosotros mismos dentro de una estructura objetiva. Dicha visión únicamente resulta posible con ayuda de un lenguaje descriptivo. (Popper, 1997, p.168)

La capacidad de articular nuestros pensamientos, reconocerlos como propios y comunicarlos nos permite, a su vez, comunicarnos con los demás, compartir ideas y construir conocimiento de manera

colaborativa. Al formular argumentos, sometemos nuestras ideas a un escrutinio crítico, lo que nos permite evaluar su validez y mejorar nuestra comprensión del mundo.

Popper afirma que alcanzamos la plena consciencia – o el estado más elevado de consciencia – cuando pensamos, especialmente cuando intentamos formular nuestros pensamientos en forma de enunciados y argumentos:

Tanto si lo hacemos en silencio al hablar con nosotros mismos – como todos lo hacemos algunas veces, a pesar del hecho de que esto ha sido negado – como si discutimos sobre algún problema interesante con nuestros amigos, o como fijamos nuestras teorías sobre el papel, no cabe ninguna duda de que el pensar y el argumentar tienen lugar en el nivel más elevado de consciencia. Al pensar articuladamente sabemos realmente que somos conscientes. (Popper, 1997, pp.167-168)

Esta capacidad de pensar articuladamente no es un logro aislado. Está íntimamente ligada al Mundo 3, el mundo de los contenidos objetivos del pensamiento. En la jerarquía de los controles, el Yo no es el centro de control más elevado, dado que a su vez está controlado plásticamente por las teorías del Mundo 3. Pero este control es, como todos los controles plásticos, de tipo «toma y daca» o de retroalimentación:

*«Esto es, podemos cambiar – y, de hecho, lo hacemos – las teorías de control del mundo 3.»* (Popper, 1997, p.169)

No somos, por tanto, simples receptores de ideas. Somos co-creadores de nuestro propio mundo de significados. Nuestro comportamiento y pensamientos se encuentran influenciados por las teorías y conocimientos que forman parte del Mundo 3, pero también podemos modificarlos. Esta relación dinámica entre el individuo y el conocimiento objetivo es, para Popper, la marca distintiva de la consciencia humana en su nivel más elevado.

En síntesis, para Popper la consciencia es un fenómeno emergente que requiere una base biológica específica, una historia evolutiva y la capacidad de generar creativamente contenidos para el Mundo 3. La memoria, el control plástico, la autoconsciencia y el lenguaje argumentativo son sus rasgos distintivos. Con estos criterios en mano,

podemos preguntarnos ahora: ¿cumple una inteligencia artificial con alguna de estas condiciones?

### **¿Puede una inteligencia artificial ser consciente? Diálogo con el debate contemporáneo**

El desarrollo vertiginoso de las inteligencias artificiales en las últimas dos décadas –especialmente los modelos de lenguaje de gran escala (LLMs) como ChatGPT, Gemini y Claude– ha reavivado con fuerza una pregunta que durante décadas parecía confinada a la ciencia ficción: ¿puede una máquina llegar a ser consciente? Este apartado aplica el marco popperiano desarrollado anteriormente al problema de la consciencia artificial. Se sostendrá que, desde la perspectiva de Popper, las IA actuales no cumplen con los criterios necesarios para ser consideradas genuinamente conscientes.

En los últimos años, el debate sobre la consciencia artificial ha cobrado centralidad. Chalmers (2023), una de las voces más influyentes, sugiere que, aunque es poco probable que los LLMs actuales sean conscientes, debemos tomar seriamente la posibilidad de que sus sucesores puedan serlo. Desde un enfoque funcionalista, Chalmers (2023) sostiene que la consciencia podría ser producto de ciertos tipos de procesamiento de información, independientemente del sustrato biológico o siliconado.

Por el contrario, el neurocientífico Seth (2021) defiende que la consciencia está profundamente ligada al hecho de que estamos vivos: tenemos cuerpos que regulan sus estados internos a través de procesos homeostáticos, y evolucionamos respondiendo al dolor y al placer con fines de supervivencia. La IA, en cambio, es un sistema de reconocimiento de patrones sobre hardware de silicio, sin metabolismo, sin cuerpo y sin historia evolutiva.

Bostrom (2014), por su parte, señala que la superinteligencia no implica necesariamente consciencia, pero advierte que es urgente comprender las condiciones bajo las cuales los «espíritus digitales» podrían tener algún grado de consciencia, pues de ello dependen importantes cuestiones morales.

Como veremos, el marco popperiano se alinea más con la posición de Seth (2021) que con la de Chalmers (2023), aunque con matices propios. A partir de lo desarrollado en los apartados anteriores, podemos sintetizar

los criterios que, desde la visión popperiana, son necesarios para la consciencia genuina:

- Base biológica y evolución: la consciencia emerge de un proceso evolutivo que ha seleccionado ventajas adaptativas en organismos vivos.
- Memoria con historia: requiere memoria como huella de una historia individual que afecta al organismo, no mero almacenamiento de datos.
- Control plástico: capacidad de generar respuestas flexibles, adaptativas y novedosas, aprendiendo de la experiencia.
- Experiencia subjetiva (Mundo 2): tener una perspectiva de primera persona, «algo que se siente ser» ese sistema.
- Autoconsciencia y creatividad argumentativa: capacidad de verse a sí mismo y de generar enunciados originales que aporten contenido nuevo al Mundo 3.

a) Base biológica y evolución. Las IA no tienen historia evolutiva. No han sido moldeadas por la selección natural en entornos de supervivencia. Son productos de diseño humano, optimizadas por algoritmos que nada tienen que ver con la lucha por la existencia que, según Popper, dio origen a la consciencia como ventaja adaptativa.

b) Memoria con historia. Popper criticaba el paralelismo psicofísico precisamente porque conducía a una «consciencia sin memoria» — meras chispas aisladas sin conexión (Popper, 1997, p.162). Las IA tienen almacenamiento de información, pero no memoria biológica. Esta última es inherentemente histórica: cada experiencia transforma al organismo de manera única. La memoria de una IA puede ser copiada o resetada sin que ello constituya una «historia».

c) Control plástico. Las IA exhiben una impresionante flexibilidad conductual. Sin embargo, esta plasticidad es simulada, no autónoma. El sistema no elige libremente entre opciones; calcula probabilidades basándose en patrones estadísticos. No hay, detrás de esa flexibilidad, un «Yo» que decida. Seth advierte que la conducta inteligente no constituye evidencia de experiencia subjetiva: un sistema puede procesar información y producir respuestas sofisticadas sin que haya nada que se sienta ser ese sistema (Seth, 2021).

d) Experiencia subjetiva (Mundo 2). Este es el criterio más difícil de evaluar, pues la experiencia subjetiva es privada. Sin embargo, desde el marco popperiano, la atribución de Mundo 2 a una entidad no biológica requiere una justificación positiva. En ausencia de base biológica compartida y de un cuerpo que siente, la carga de la prueba recae sobre quienes afirman la consciencia artificial. Hasta ahora, no existe evidencia concluyente de que las IA tengan experiencias subjetivas.

e) Autoconsciencia y creatividad. Popper vinculaba la autoconsciencia a la capacidad de «verse a sí mismo desde el exterior» mediante el lenguaje descriptivo (Popper, 1997, p.168). Las IA pueden producir enunciados como «yo creo que...», pero esto es un artefacto lingüístico, no evidencia de un Yo sustantivo. Para comprender por qué, conviene precisar qué entiende Popper por creatividad genuina. En su sentido popperiano, la creatividad no es mera novedad combinatoria, sino la capacidad de un sujeto del Mundo 2 de generar contenidos que, siendo originados en su experiencia subjetiva, trascienden los datos disponibles y aportan estructuras argumentativas inéditas al Mundo 3. Esta capacidad presupone intencionalidad —el pensamiento está dirigido hacia algo— y reflexividad —el sujeto sabe que está pensando y puede evaluar críticamente sus propios argumentos—.

Los LLMs, en cambio, operan mediante predicción estadística de secuencias: dado un contexto, calculan la continuación más probable según patrones aprendidos de grandes corpus textuales. Su «novedad» es combinatoria —reorganización de estructuras previas— y no implica ningún Yo que evalúe, corrija o se comprometa con lo producido. Una distinción útil aquí es la que traza Searle (1980) con su Argumento del Cuarto Chino: la manipulación sintáctica de símbolos no equivale a comprensión semántica. Ahora bien, el marco popperiano no se limita a suscribir esta crítica, sino que la amplía en una dirección propia. Mientras Searle se apoya en la semántica —la comprensión del significado— para negar que la sintaxis baste para la mente, Popper añade dos dimensiones adicionales: la experiencia subjetiva del Mundo 2 —que algo se sienta ser ese sistema— y la creación de contenido objetivo para el Mundo 3 —que el sistema aporte argumentos originales que trasciendan los datos con los que fue entrenado—. Dicho de otro modo: aunque un LLM manipulara símbolos de forma semánticamente competente, seguiría sin satisfacer las condiciones popperianas para la consciencia. Tres razones sintetizan la postura:

1. La consciencia es un fenómeno emergente ligado a la evolución biológica. Las IA no tienen historia evolutiva ni luchan por sobrevivir en entornos que recompensan la adaptación consciente.
2. La consciencia requiere un cuerpo y experiencia subjetiva. Como argumenta Seth, la consciencia biológica está ligada a la regulación homeostática, al dolor y al placer. Las IA no tienen cuerpo, no sienten, no tienen un «bienestar» que proteger.
3. La consciencia requiere creatividad genuina, no mera recombinación estadística. Las IA son sistemas predictivos que recombinan patrones aprendidos. No crean argumentos originales que aporten contenido nuevo al Mundo 3. Son, como ya se anticipó, productos del Mundo 3, no habitantes del Mundo 2.

Esto no significa que las IA sean irrelevantes para la filosofía de la mente. Todo lo contrario. Su desarrollo nos exige repensar nociones tradicionales (Martins-Costa y Prazeres-Martins, 2025). Pero esa reflexión no debería apresurarse a atribuir consciencia donde solo hay simulación. La visión popperiana nos invita a considerar las inteligencias artificiales como herramientas poderosas que amplían nuestras capacidades cognitivas, pero no como sujetos conscientes. La consciencia, al menos por ahora, sigue siendo un fenómeno irreductiblemente biológico.

## Discusión

La postura popperiana defendida en este artículo puede situarse en el mapa del debate contemporáneo sobre consciencia e inteligencia artificial.

Frente al funcionalismo computacional (Chalmers, 2023), Popper diverge profundamente. Para esta corriente (Chalmers, 2023; Putnam, 1967), la consciencia depende de la organización funcional del sistema —los patrones de procesamiento de información— y no del sustrato en que se implementa. Si el cerebro humano produce consciencia en virtud de su arquitectura funcional, entonces un sistema artificial con arquitectura funcionalmente equivalente debería producirla también. Esta es la intuición de fondo detrás de la hipótesis de que los LLMs avanzados podrían ser conscientes.

Popper rechaza esta tesis por una razón que va más allá de la mera preferencia por el sustrato biológico. Para él, la consciencia no es un patrón funcional abstracto, sino un fenómeno que emerge de una historia: la historia evolutiva de la especie y la historia individual del organismo. Esta historia no es un detalle accidental – es constitutiva. Un sistema que no ha sido moldeado por la presión selectiva de la supervivencia, que no ha experimentado dolor ni placer como señales con consecuencias reales, y que no posee una memoria que lo transforme acumulativamente como individuo, no cumple las condiciones bajo las cuales emerge la consciencia. El funcionalismo, al abstraer la consciencia de su historia, comete el error que Popper identificaba en el paralelismo psicofísico: postula una «consciencia sin memoria», es decir, sin la continuidad que hace posible la experiencia subjetiva como tal.

Dicho de otro modo, el error del funcionalismo no es afirmar que el sustrato es irrelevante en principio, sino ignorar que el sustrato biológico no es un mero recipiente, sino el medio en el que la consciencia se constituye históricamente. Desde esta perspectiva, la pregunta pertinente no es «¿puede este sistema procesar información de manera funcionalmente equivalente al cerebro?», sino «¿tiene este sistema la historia que hace posible que emerja algo que se sienta ser ese sistema?». Las IA actuales, por diseño, no la tienen.

Frente al enactivismo (Seth, 2021; Thompson, 2010), en cambio, hay convergencia importante. Ambas posturas anclan la consciencia en la interacción corporizada del organismo con su entorno. Sin embargo, Popper otorga al Mundo 3 un peso explicativo que las posturas enactivistas no suelen desarrollar en la misma dirección – el conocimiento objetivo, las teorías – como producto y condición de la consciencia humana en su nivel más elevado.

Toda postura filosófica tiene limitaciones, y la de Popper no es la excepción. En primer lugar, Popper no ofrece un criterio positivo para determinar qué organismos tienen consciencia y en qué grado. Señala grados, pero no umbrales. En segundo lugar, su tesis de que la consciencia requiere necesariamente un sustrato biológico es una afirmación empírica que, en principio, podría ser falsada si en el futuro se desarrollara un sistema artificial con memoria con historia, control plástico autónomo y creatividad genuina.

A pesar de estas limitaciones, la visión popperiana ofrece una contribución valiosa: en un contexto donde se tiende a antropomorfizar los modelos de lenguaje, Popper nos recuerda la importancia de no confundir simulación con posesión. Un sistema puede comportarse «como si» tuviera consciencia sin tenerla. El marco popperiano, con su énfasis en la evolución biológica, la memoria con historia y el Mundo 3, proporciona una base sólida para iniciar esa discusión.

### Conclusiones

Este artículo parte de una pregunta central: ¿puede una inteligencia artificial ser consciente? Desde el marco popperiano desarrollado a lo largo de estas páginas, la respuesta es negativa para las IA actuales y previsiblemente para cualquier sistema que no cumpla con las condiciones biológicas y evolutivas que Popper considera constitutivas de la consciencia.

A lo largo del recorrido, se ha argumentado que la consciencia no es un mero procesamiento de información ni una sustancia separada, sino un fenómeno emergente ligado a la evolución biológica. Popper, desde su ontología de los tres mundos y su rechazo al reduccionismo materialista, sitúa la consciencia en el Mundo 2 (experiencia subjetiva) y la vincula a la memoria con historia, al control plástico y a la capacidad creativa de generar argumentos originales que aportan contenido nuevo al Mundo 3. Las inteligencias artificiales, por el contrario, son productos del Mundo 3 que simulan procesos cognitivos sin la base biológica necesaria, sin experiencia subjetiva y sin creatividad genuina – solo recombinación estadística de datos previos.

Como se discutió, esta postura no es incompatible con reconocer las limitaciones del enfoque popperiano – especialmente la ausencia de un criterio positivo para determinar grados de consciencia – ni con estar abiertos a futuras revisiones empíricas. Sin embargo, ofrece una advertencia metodológica relevante en el contexto actual: no debemos confundir la simulación de consciencia con su posesión.

Futuras investigaciones podrían explorar si algunos de los criterios popperianos – como el control plástico o la memoria con historia – pueden ser operacionalizados para su evaluación en sistemas artificiales, así como dialogar con propuestas más recientes que buscan extender la ontología popperiana al dominio digital. Mientras tanto, la consciencia

sigue siendo, desde esta perspectiva, un fenómeno irreductiblemente biológico.

## REFERENCIAS

Austin, J. L. (1989). Otras mentes. En J. O. Urmson y A. García Suárez (Eds. y Trad.), *Ensayos filosóficos* (pp. 87-117). Alianza Editorial. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press. Chalmers, D. J. (2023). Could a large language model be conscious? *arXiv*. <https://arxiv.org/abs/2303.07103>

Martins-Costa, A. y Prazeres-Martins, F. (2025). Conciencia e inteligencia en la era de los grandes modelos de lenguaje. *Resonancias. Revista de Filosofía*, (20), 83-93. <https://doi.org/10.5354/0719-790X.2025.80188>

Popper, K. (1992). *Conocimiento objetivo: Un enfoque evolucionista* (4.<sup>a</sup> ed., C. Solís Santos, Trad.). Tecnos. [https://padron.entretemas.com.ve/cursos/Epistem/Libros/Popper\\_ConocimientoObjetivo.pdf](https://padron.entretemas.com.ve/cursos/Epistem/Libros/Popper_ConocimientoObjetivo.pdf)

Popper, K. (1997). *El cuerpo y la mente*. Paidós. <https://tavapy.gov.py/biblioteca/wp-content/uploads/2023/01/PopperK-El-cuerpo-y-la-mente.pdf>

Putnam, H. (1967). Psychological predicates. En W. H. Capitan y D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37-48). University of Pittsburgh Press. <https://www.phil.uu.nl/~joel/3027/3027PutnamPsychPredicates.pdf>

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-457. <https://doi.org/10.1017/S0140525X00005756>

Seth, A. (2021). *Being you: A new science of consciousness*. Faber y Faber. [https://files.spiritmaji.com/books/self-help/Being%20You%20\\_%20A%20New%20Science%20of%20Co%20-%20Seth,%20Anil.pdf](https://files.spiritmaji.com/books/self-help/Being%20You%20_%20A%20New%20Science%20of%20Co%20-%20Seth,%20Anil.pdf)

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*. Belknap Press.

### Cómo citar:

Valdes-Millan, A. (2025). ¿Puede una máquina ser consciente? Una respuesta popperiana al problema de las otras mentes en la era de la inteligencia artificial. *Revista Discusiones Filosóficas*. 26(47), 95-112. <https://doi.org/10.17151/difil.2025.26.47.6>