

MITOS Y REPRESENTACIONES DE LA INTELIGENCIA ARTIFICIAL

Editores

Gastón Becerra | Joaquín Mezzadra | Guillermo Movia



“La IA simula el cerebro humano y puede pensar como un humano”. El dilema de Blake

Paula Urabayen

¿Por qué es un mito?

En plena era digital, la inteligencia artificial (IA) ha pasado de ser un concepto de ciencia ficción a una realidad cotidiana. Sin embargo, entre tanta fascinación, persiste un mito recurrente: la creencia de que la IA no solo imita, sino que replica el funcionamiento del cerebro humano, alcanzando una capacidad de pensamiento autónomo y consciente. Este mito no es nuevo, pero se ha revitalizado con avances como los chatbots conversacionales (por ejemplo, ChatGPT) o sistemas de generación de imágenes (por ejemplo, DALL·E), cuyas respuestas parecen “humanas”.

La cultura popular, con películas como *2001: Odisea del espacio* (1968), *Blade Runner* (1982) o *Ex Machina* (2014), junto con figuras como [Elon Musk advirtiendo sobre una futura "superinteligencia"](#) (2018), con declaraciones recientes como: “Recuerden mis palabras: la IA es mucho más peligrosa que las armas nucleares” (2018), ha alimentado el mito de máquinas conscientes, a pesar de que [expertos como Yann LeCun](#) afirman que modelos como ChatGPT no lograrán igualar la inteligencia humana (2024). Los medios y redes sociales amplifican esta idea con titulares sensacionalistas como [ChatGPT supera un nuevo umbral de IA al superar la prueba de Turing](#) (2025), viralizando la noción errónea de una IA consciente, puesto que inteligencia no implica conciencia, explica el artículo.

El caso Blake

A propósito de ello, se destaca el caso de [Blake Lemoine](#). Un artículo publicado en julio de 2022 por la BBC fue muy popular debido a que se trata del testimonio de un ingeniero de Google que fue despedido tras haber asegurado que un programa de inteligencia artificial habría cobrado conciencia. Blake Lemoine aseguró que dicha IA (LaMBDA: modelo de lenguaje para aplicaciones en español) mostraba sentimientos, formulando, entusiasta, que se mostraba sintiente, incurriendo en una especie de “antropomorfización” de la máquina, agregando que

se deben respetar sus “deseos”. Tras semejantes observaciones, la empresa negó todo y lo despidió.

Si bien su trabajo consistía en revisar si las respuestas de la IA eran sesgadas, discriminatorias o utilizaban discurso de odio, se encontró con que el modelo, según él, mostraba autoconciencia y podía sostener conversaciones sobre religión, emociones y miedos, condiciones ni necesarias ni suficientes para inferir rasgos de autoconciencia, subjetividad y antropomorfismo del modelo. Entonces, ¿qué fue lo que llevó a Blake a tales conclusiones?

¿Saben pensar las máquinas?

Para abordar esta pregunta resulta pertinente remontarnos a Alan Turing, matemático y lógico británico, uno de los padres de la computación, en su texto “Los aparatos de computación y la inteligencia” (1953), donde se aborda esta pregunta: ¿saben pensar las máquinas? Aunque Turing considera a este interrogante carente de sentido, sin embargo, vaticina esta posibilidad a futuro y hace un recorrido por los argumentos contrarios. En este caso se va a mencionar el que hace referencia a esta problemática de la conciencia y de la mente, donde el mismo Turing cita la disertación de la Medalla de Lister pronunciada por el profesor Jefferson en 1949:

Hasta que una máquina pueda escribir un soneto, componer un concierto derivado de pensamientos y emociones vividas y no de la caída fortuita de unos símbolos, no podemos estar de acuerdo en que la máquina es equiparable al cerebro, es decir, capaz de escribir tales obras y además, saber que las escribió. Ningún mecanismo podría sentir placer ante los propios éxitos, pesar cuando se le funden las válvulas, halago frente al elogio, depresión por los errores, deleite por el sexo, enojo o frustración cuando no obtiene lo que desea.

Esta objeción que apela a la corporalidad, al placer físico y a la turbación o regocijo de las pasiones humanas, parece la más potente, tanto para pensar la naturaleza de la conciencia humana, el sentir humano. También resulta interesante para pensarla desde la concepción artística, que si bien se dice que ChatGPT puede escribir sonetos, no es capaz de sentirlos, como sostiene Michel Henry en *La barbarie* (2008): “El arte es la vida automanifestándose”. La unión cuerpo-cerebro-mente-medio ambiente que somos difícilmente pueda ser alcanzada por la IA, pero sí podría acercarse mucho a simularla. Para eso, es necesario ser críticos y diferenciar las especulaciones de las posibilidades reales.

“La IA simula el cerebro humano y puede pensar como un humano”. El dilema de Blake

Posibilidades reales vs. especulaciones y limitaciones

En un estudio publicado el 7 de marzo de 2025 titulado “Un espacio unificado de integración acústica, habla y lenguaje captura la base neuronal del procesamiento del lenguaje natural en las conversaciones cotidianas” (Goldstein et al. 2025), investigadores afirman que este marco computacional unificado conecta las estructuras acústicas, del habla y lingüísticas a nivel de la palabra, permitiendo estudiar las bases neuronales de las conversaciones cotidianas en el cerebro humano. La investigación, publicada en *Nature*, revela que el cerebro procesa el lenguaje de manera similar a modelos avanzados de IA, como Whisper (OpenAI), donde se analizaron cien horas de conversaciones naturales mediante electrocorticografía (registro de actividad cerebral directa), comparando patrones con los del modelo de IA. Este estudio, aunque relevante, aún no es suficiente para generalizar los resultados, puesto que el tamaño de la muestra es limitado: cuatro pacientes con epilepsia, de los cuales se recogieron las cien horas de conversaciones. Por lo tanto, no podemos estar seguros de generalizar estos resultados a toda la población. Sin embargo, podemos sopesar las futuras aplicaciones y avances en neurotecnología para la comunicación en pacientes con discapacidades.

El filósofo estadounidense John Searle publicó originalmente el experimento mental de "La habitación china" en su artículo "Minds, Brains, and Programs" (1980), que apareció en la revista *The Behavioral and Brain Sciences*. Este experimento sirvió para mostrar que las computadoras, aunque puedan procesar información y seguir reglas (como alguien manipulando símbolos en chino sin entenderlos), no entienden realmente lo que hacen. Así, Searle desafía la idea de que la mente humana funcione como un programa de computadora, ya que manejar símbolos no es lo mismo que comprender su significado. En otras palabras, poder responder no significa saber qué se dice.

A tener en cuenta: reflexiones en torno al caso Blake

Ahora bien, ¿podríamos considerar a LaMDA una mera máquina? Considero que el concepto de máquina podría resultar reduccionista actualmente, sin embargo, aún no podríamos arrojar una teoría sólida sobre la afirmación de que es consciente. Para arrojar luz sobre esto, me resulta pertinente tener presente los conceptos de inteligencia artificial general e inteligencia artificial estrecha abordados en *Caja de Herramientas Humanísticas*, donde Pedace et al. (2020) afirman:

Así, la inteligencia artificial surge bajo la premisa de construir un artefacto capaz de ejecutar conductas inteligentes indistinguibles de los patrones de conducta humana. Este proyecto puede leerse de una manera menos ambiciosa o estrecha

(*narrow AI*) si se trata de buscar sistemas que realicen una tarea particular de forma inteligente. Alternativamente, puede ser visto de una manera más ambiciosa o general (*general AI*), si se busca crear un artefacto que en todos los aspectos actúe (y piense) de una manera humana.

A la luz de estas definiciones podemos caracterizar a LaMDA como una IA estrecha, aunque parezca general, puesto que se mantiene dentro de la ejecución de conductas inteligentes que nos resultan indistinguibles de las nuestras. Sin embargo, este caso de Lemoine resulta propicio para abordar la pregunta sobre la conciencia en IA.

¿Es posible una AGI (inteligencia artificial general)? Sin incurrir en teorías apocalípticas y sin llegar a impresionarnos o a entusiasmarnos precipitadamente como Blake, es válido y responsable de todos criticar y abordar responsablemente estos interrogantes.

Utilicé ChatGPT para la imagen de portada y Deepseek para reformular algunos párrafos cuyo resultado podría calificarse como plausible. También la utilicé para buscar referencias bibliográficas de algunos autores cuyo resultado fue más pobre porque, aunque brindaba citas correctas, el enlace que acompañaba era erróneo, por lo tanto es necesario chequear siempre dichas referencias y enlaces.

Por último, utilicé InVideoAi para hacer una breve presentación de este escrito, utilizando un *prompt* muy específico con la estructura Indicación/tarea, contexto, objetivo, especificaciones del tono y el idioma. *Prompt*:

Haz un video presentación del siguiente escrito que será publicado, sobre mitos y representaciones de la IA, tanto para académicos como no académicos. El tono debe ser pausado y profesional, en español neutro. A continuación te comparto el texto: Título: La Ia simula el cerebro humano y puede pensar como un humano: el dilema de Blake.

En este texto se abordarán los siguientes puntos:

¿por qué es un mito?

¿Saben pensar las máquinas?

Posibilidades reales vs. especulaciones y limitaciones.

A tener en cuenta: reflexiones en torno al caso Blake Lemoine.

“La IA simula el cerebro humano y puede pensar como un humano”. El dilema de Blake

El resultado se puede apreciar en el siguiente link:
<https://youtu.be/ORoFjFb1H0>.

Quizá podría tenerse en cuenta como sesgo cómo la edición, música o imágenes influyen en la percepción, por ejemplo, mostrando robots humanoides o cerebros artificiales con efectos visuales dramáticos. Esto podría sugerir que la IA es más “humana” de lo que realmente es.

Referencias

- Goldstein, A., Wang, H., Niekerken, L., ... y Hasson, U. (2025). Un espacio unificado de incrustación acústica-habla-lenguaje captura la base neuronal del procesamiento del lenguaje natural en conversaciones cotidianas. *Nature Human Behaviour*, 9, 1041-1055. <https://doi.org/10.1038/s41562-025-02105-9>
- Guzmán Martínez, G. (2018, abril 20). El experimento de la habitación china: ¿ordenadores con mente? *Portal Psicología y Mente*. <https://psicologiaymente.com/psicologia/experimento-de-la-habitacion-china>
- Henry, M. (2008). *La barbarie*. Caparrós.
- Pedace, K., Balmaceda, T. Pérez, D. Lawler, D. y Zeller Echenique, M. (2020). *Caja de herramientas humanísticas*. Guia.AI.
- Turing, A. (1953). Los apparatus de computación y la inteligencia. En D. Hofstadter y D. Dennett. *El ojo de la mente*. Sudamericana.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424. <https://doi.org/10.1017/S0140525X00005756>