¿Estamos preparados para detectar la consciencia en las máquinas?

La pregunta de si las máquinas podrían desarrollar consciencia es una pregunta que nos venimos planteando desde hace muchas décadas, que ha recobrado un gran interés debido a estos desarrollos tecnológicos. Más allá de si las máquinas podrían desarrollar consciencia, deberíamos preguntarnos si la comunidad científica estaría preparada para detectar consciencia en las máquinas si ésta llegase a surgir.

[ILUSTRACIÓN: -IZABELL-/ ISTOCK]

El desarrollo de la inteligencia artificial ha sido vertiginoso en los últimos años. La aparición de modelos de inteligencia artificial de acceso al público general como ChatGPT nos sorprendió por la capacidad de generar contenido, comprender los mensajes del usuario y responder de forma claramente estructurada (y con acierto, en la mayoría de las ocasiones). Pero, ¿implica eso que tienen consciencia? Lejos de ser nueva, venimos planteando esta pregunta desde hace muchas décadas. Más allá de si las máquinas podrían desarrollar consciencia, deberíamos preguntarnos si la comunidad científica estaría preparada para detectar consciencia en las máquinas si ésta llegase a surgir.

No existe una definición consensuada de consciencia

En primer lugar, para poder detectar consciencia debemos contar con una definición consensuada del término. A día de hoy, esta definición no existe.

Los humanos (y otras especies animales) somos capaces de procesar una parte limitada de la información que entra en nuestros sentidos, y sólo podemos reportar una parte aún más limitada. Si vemos, por ejemplo, el cuadro de *Las Meninas* durante un segundo, podríamos reportar a las meninas y a algunos otros personajes que se encuentran en una habitación, pero no podríamos reportar el espejo, el perro o algunos otros detalles de la pintura. Según algunos autores, la información que podemos reportar representa nuestra experiencia consciente.

Otros autores piensan que nuestra consciencia es más rica que la información que podemos reportar. En el caso del cuadro de Las Meninas, se plantea que nuestra consciencia abarca prácticamente todos los aspectos del cuadro, aunque no podamos reportarlos. Esto es lo que se conoce como consciencia fenomenológica, que se refiere a la experiencia subjetiva de la conciencia. El estudio de la consciencia fenomenológica intentaría responder a cuestiones como qué siente un humano cuando ve el color rojo.

La experiencia fenomenológica puede medirse con paradigmas experimentales que nos indican si las respuestas o las reacciones fisiológicas son iguales o distintas ante estimulación diferente. Es decir, podríamos medir si una abeja responde de manera distinta al rojo o al azul, aunque es más difícil intentar comprender qué siente una abeja cuando percibe esos colores.

Podríamos medir si una abeja responde de manera distinta al rojo o al azul, pero nos costaría comprender qué siente una abeja cuando percibe esos colores

En humanos y otros animales podemos medir la consciencia estudiando las respuestas de los organismos. Sus respuestas verbales o no verbales, las respuestas de su cerebro o de su corazón.

Cuando una persona está en coma, medimos su consciencia a través de la respuesta a la estimulación exterior. Cuánto más elaborada la respuesta, mayor consciencia. Así, un paciente que responda al dolor deberá tener cierto grado de consciencia. Si responde a órdenes verbales simples, por ejemplo, a la orden "levanta un dedo", su consciencia será mayor.

Usando técnicas de imagen cerebral también puede estudiarse la respuesta cerebral. La respuesta de ciertas regiones cerebrales se ha asociado a mayores niveles de consciencia, de acuerdo a todo el conocimiento acumulado experimentalmente sobre cómo responde el cerebro cuando percibimos conscientemente.

¿Cómo estudiar la consciencia en un organismo no biológico?

Según algunos autores, la consciencia es una computación, de manera que si un sistema realiza los cómputos necesarios, podemos decir que es consciente. Pongamos como ejemplo los cómputos que plantea la teoría del Espacio de Trabajo Global (Dehaene et al., 2017). Esta teoría distingue tres tipos diferentes de computaciones: C0, que se refiere a computaciones no conscientes (es decir, operaciones de procesamiento de información que pueden llevarse a cabo sin consciencia), característica de muchos algoritmos actuales. C1 representa el nivel de acceso consciente, donde un espacio de trabajo transmite y amplifica globalmente una pieza específica de información, similar a un sistema de enrutamiento. Es decir, la información consciente se amplifica y es accesible por diferentes sistemas, por ejemplo, podemos responder a ella de manera verbal y con una respuesta motora. C2 se refiere a la capacidad de un sistema para representarse a sí mismo. Un sistema que posee tanto autorrepresentación como acceso consciente a esa representación tendrá conocimiento sobre sí mismo: sabrá lo que sabe y lo que no sabe. De acuerdo con estos autores, si implementamos en una máquina C1 y C2 la máquina será consciente. En el caso de un coche, éste tendría signos de consciencia si la información de los diferentes sistemas (tanque de gasolina, GPS, gasolineras más cercanas, información sobre averías, talleres más cercanos, etc.) se comparte en un espacio de trabajo global que haga que la información de uno de los sistemas (me estoy quedando sin gasolina) esté disponible en los demás sistemas (dónde está la próxima gasolinera). Por otro lado, la máquina debería crear una autorrepresentación, teniendo conocimiento sobre sí misma.

La consciencia es una computación, de manera que si un sistema realiza los cómputos necesarios, podemos decir que es consciente

Otros autores tienen teorías sobre la consciencia basadas en la biología del cerebro animal. Por ejemplo, la teoría del procesamiento recurrente plantea que el procesamiento recurrente, que envía información arriba y abajo en el sistema, en determinadas regiones cerebrales está ligado a la consciencia. Sin embargo, el procesamiento recurrente en otros sistemas no biológicos tiene necesariamente características diferentes, que no tienen por qué asociarse con el procesamiento consciente.

Para otros autores, implementar estos cómputos en una máquina no sería suficiente para poder afirmar que la máquina es consciente. Se plantea que la consciencia está ligada a la existencia de un sistema nervioso, a su relación con el organismo y a la necesidad de supervivencia. Sin un organismo vivo, no se puede generar consciencia, ya que la consciencia está intrínsecamente ligada a nuestras emociones y a la necesidad de mantener a nuestro organismo con vida. Entonces la pregunta es: ¿podría desarrollar una máquina consciencia si es programada para sentir emociones e intentar protegerse? A día de hoy no existen evidencias de aspectos emocionales en máquinas y la comunidad científica está reflexionando sobre la conveniencia de programar máquinas pudieran sentir emociones o que tengan la finalidad de protegerse.

Sin un organismo vivo, no se puede generar consciencia, ya que la consciencia está intrínsecamente ligada a nuestras emociones y a la necesidad de mantener a nuestro organismo con vida

En conclusión, la respuesta a nuestra pregunta "¿estamos preparados para detectar la consciencia en las máquinas?" es no. Nuestras teorías se han desarrollado para explicar la consciencia en organismos vivos y al no contar con una definición de consciencia, hoy en día sería teóricamente posible que una máquina la genere y no podamos detectarlo.

Block, N. *Perceptual consciousness overflows cognitive access* en Trends in Cognitive Sciences. (2011), *15*(12), pp: 567–575). Disponible en: https://doi.org/10.1016/j.tics.2011.11.001

Damasio, A., & Damasio, H. *Feelings are the source of consciousness* en Neural Computation. (2023, *35*(3), pp: 277–286). Disponible en: https://doi.org/10.1162/neco_a_01521

Dehaene, S., Lau, H., & Kouider, S.. *What is consciousness, and could machines have it?* en Science. (2017, *358*(6362), pp: 486–492). Disponible en: https://doi.org/10.1126/science.aan8871

Lau, H. (2022): *In Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience*. Oxford, Oxford University Press. Disponible en: https://global.oup.com/academic/product/in-consciousness-we-trust-9780198856771

Mudrik, L., Boly, M., Dehaene, S., Fleming, S. M., Lamme, V., Seth, A., Melloni, L. *Unpacking the complexities of consciousness: Theories and reflections* en Neuroscience & Biobehavioral Reviews (2025; 170: 106053). Disponible en: https://www.sciencedirect.com/science/article/pii/S0149763425000533