Cuadernos de Neuropsicología / Panamerican Journal of Neuropsychology
2024, Vol. 18 Nº1 07 - 08
ISSN: 0718-4123
DOI: 10.7714/CNPS/18.1.101

Editorial. ¿La Inteligencia Artificial puede desarrollar consciencia?

Can Artificial Intelligence Develop Consciousness?

A Inteligência Artificial pode desenvolver consciência?

Marcos Domic-Siede[a]

a. Laboratorio de Neurociencia Cognitiva, Escuela de Psicología, Universidad Católica del Norte.

Cuando Blake Lemoine, ingeniero de Google, declaró que LaMDA, el chatbot en el que había estado trabajando, poseía consciencia, fue posteriormente despedido de la empresa. Este comentario avivó un debate en curso entre expertos en tecnología, filósofos y personalidades públicas sobre la posibilidad de que la inteligencia artificial (IA) alcance algún grado de consciencia en el futuro. Sistemas de IA como LaMDA y ChatGPT pueden dar la impresión de tener consciencia. Sin embargo, se les ha entrenado utilizando enormes volúmenes de texto, por lo que han aprendido a responder de forma similar a como lo haría un humano. Así que, ¿cómo podríamos estar seguros si tienen o podrían desarrollar una forma de consciencia? Hay varios aspectos que debemos tener en cuenta. El primero es ¿Qué entendemos por consciencia? Si decimos que es sinónimo de inteligencia, entonces los sistemas de IA serían conscientes. Pero inteligencia no es sinónimo de consciencia. Cuando miramos el mundo animal, nos encontramos con formas de vida que exhiben un sistema cognitivo complejo, como algunos córvidos, delfines y primates, mostrando ser capaces de resolver problemas, tomar decisiones, y una comunicación sofisticada. Sin embargo, dado que los animales no pueden realizar reportes verbales, es más desafiante poder determinar con seguridad sobre si su comportamiento es controlado conscientemente.

Pero ¿Qué es la consciencia? Tradicionalmente, la consciencia se ha abordado desde una perspectiva dicotómica, como un estado unitario de consciencia/inconsciencia. Sin embargo, Endel Tulving, un destacado psicólogo experimental y cognitivo, introdujo una visión más matizada con su taxonomía tripartita de la consciencia: autonoética, noética y anoética. La consciencia autonoética se relaciona con la capacidad de revivir eventos pasados y anticipar futuros, implicando un sentido del "yo" a través del tiempo, estrechamente ligada a nuestra memoria autobiográfica. La noética, asociada a la memoria semántica, abarca el conocimiento del mundo y los hechos, independientemente de la experiencia personal. Por otro lado, la consciencia anoética, relacionada con memorias procedimentales y respuestas emocionales o fisiológicas evocadas sin conocimiento consciente, opera en un nivel subterráneo, guiándonos más por sensación que por razonamiento explícito, como el sentir una inquietud inexplicable o reconocer instintivamente un objeto en nuestro campo visual.

En el espectro humano de la consciencia, la "consciencia anoética" actúa como sustrato basal para experiencias más complejas y autorreflexivas. Si consideramos otros animales, como perros o ratones, su experiencia del mundo podría predominar en la consciencia anoética, basada en sensaciones inmediatas de confort o peligro. Los humanos, en cambio, representamos constantemente nuestras propias operaciones cognitivas, una "consciencia de la consciencia" donde lo básico es captado y modulado por niveles cognitivos superiores. Este es un tema fascinante y debatido, al igual que la pregunta de qué animales más allá de los humanos experimentan diferentes tipos de consciencia. Mientras la "consciencia autonoética" y la narrativa extendida en el tiempo podrían ser exclusivas de los humanos y algunos primates, la "consciencia noética" y la consciencia anoética podrían ser más universales, compartidas por una amplia gama de animales, sugiriendo que la sensibilidad o sintiencia es una forma de consciencia anoética que, aunque carece de un 'yo' narrativo, es capaz de experimentar directamente.

Correspondencia: Marco Domic-Siede. Escuela de Psicología, Universidad Católica del Norte. E-mail: mdomic@msn.com

ISSN: 0718-4123
DOI: 10.7714/CNPS/18.1.101

¿Podría un algoritmo de IA alguna vez tener sensibilidad o sintiencia? Primero, debemos cuestionar nuestras premisas y reconocer la posibilidad de estar seducidos por una "ilusión de similitud". La tendencia humana a antropomorfizar, como se evidenció en el experimento de Heider y Simmel de 1944, nos lleva a atribuir características humanas incluso a formas geométricas simples. Así, aunque las IA puedan comunicarse de forma que recuerda a la interacción humana, esta similitud es superficial y engañosa. Los organismos biológicos, a diferencia de las IA, tienen vidas sociales complejas y una consciencia moldeada por intercambios sociales y culturales, mientras que las IA, carentes de cultura propia o diversidad genética, son meras extensiones de objetivos humanos, programadas para tareas específicas. Además, la rica y multifacética percepción sensorial y cognitiva de los organismos biológicos contrasta con la experiencia unidimensional y mediada digitalmente de la IA. La vitalidad intrínseca, la autopreservación y la adaptación activa de los sistemas biológicos son ajenas a las IA, que no evolucionan ni se adaptan más allá de su programación. Considerando la complejidad evolutiva que culminó en sistemas nerviosos que nos permiten ser conscientes, se comprende que la consciencia es el resultado de una larga historia evolutiva, una serie de capas de complejidad que permiten no solo la percepción, sino la reflexión sobre esa percepción y la metacognición. Esta capacidad de reflexionar sobre el propio pensamiento, particularmente desarrollada en humanos, nos permite no solo experimentar sensaciones, sino también discutirlas y utilizar ese conocimiento para adaptarnos de manera sofisticada a nuestro entorno, una profundidad de experiencia que está más allá del alcance actual de la IA.

Al examinar la inteligencia artificial y modelos como el chat GPT-4, nos enfrentamos a una paradoja: aunque capaces de procesar y generar lenguaje de forma avanzada, estas máquinas operan sin una conexión con la experiencia sensorial y subjetiva, careciendo de un modelo interno del mundo y de sensaciones fundamentales. Su habilidad lingüística, fruto de algoritmos avanzados, no se equipara a la evolución biológica que confiere la experiencia subjetiva. Esta falta de consciencia, un fenómeno más allá de la mera manipulación de símbolos lingüísticos es el producto de un largo proceso evolutivo que engendra experiencias subjetivas y reflexivas en organismos. La capacidad de la IA para emular la comunicación humana no debe ser malinterpretada como un indicativo de consciencia, la cual requiere atributos como sociabilidad, percepción sensorial enriquecida y vitalidad. El debate sobre la consciencia en la IA, por tanto, refleja más nuestros prejuicios y deseos antropocéntricos que las verdaderas capacidades de estas impresionantes, pero intrínsecamente limitadas máquinas.

Lecturas sugeridas

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

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

Gazzaniga, M. S. (2006). *The ethical brain: The science of our moral dilemmas* (Reprint ed.). Ecco.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243–259. https://doi.org/10.2307/1416950

Koch, C., & Tononi, G. (2019). Can machines be conscious? *IEEE Spectrum, 45*(6), 55–59. https://doi.org/10.1109/MSPEC.2008.4531463

Krauss, P., & Maier, A. (2020). Will We Ever Have Conscious Machines?. *Frontiers in computational neuroscience*, *14*, 556544. https://doi.org/10.3389/fncom.2020.556544

Nicolelis, M. (2020). *The True Creator of Everything: How the Human Brain Shaped the Universe as We Know It*. Yale University Press. https://doi.org/10.2307/j.ctvt1sggf

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1–12. https://doi.org/10.1037/h0080017

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433