La inteligencia artificial y el «problema difícil» de la conciencia

Las inteligencias artificiales, incluso las corpóreas, al no tener conciencia por el hecho de no ser seres vivos, no pueden comprender realmente ni el mundo ni el lenguaje, por mucho que nuestra tendencia a proyectar características humanas donde no las hay nos haga pensar que sí.

[ILUSTRACIÓN: YAKOVLIEV / ISTOCK]

¿Por qué los humanos, a pesar de compartir una biología similar, tenemos percepciones del mundo tan diferentes? Este enigma, que se conoce como la "brecha explicativa" o, en palabras de David Chalmers, «el problema difícil» de la conciencia, ha generado numerosos debates y especulaciones. La inmensa mayoría de los neurocientíficos piensan que el cerebro es la única causa de la mente. Pero a pesar de los avances en neurociencia y psicología, aún estamos muy lejos de cerrar la brecha explicativa.

Desde tiempos inmemoriales, la conciencia ha sido uno de los mayores misterios que ha cautivado a filósofos, científicos y pensadores. ¿Cómo es posible que procesos electroquímicos que tienen lugar en nuestro cerebro generen experiencias subjetivas?

El enfoque pragmático: parte por parte

Neurocientíficos volcados en el tema, como Anil Seth, profesor de neurociencia cognitiva y computacional en la Universidad de Sussex (Reino Unido), proponen un enfoque pragmático: en lugar de intentar resolver el problema en su totalidad, se centran en entender cómo ciertos procesos cerebrales específicos generan experiencias conscientes particulares.

De hecho, prácticamente todos los experimentos sobre la conciencia tienen en común que buscan, y a menudo encuentran, correlatos neuronales de la conciencia. Pero los correlatos no explican las preguntas de por qué y cómo la actividad física del cerebro da lugar a una experiencia subjetiva completa. Para explicarlo se necesita encontrar la cadena de relaciones causa-efecto que vinculan la actividad neuronal con la conciencia y otros procesos cognitivos de alto nivel.

Este enfoque pragmático define la conciencia, de manera minimalista, como "cualquier tipo de experiencia subjetiva, desde el placer hasta el dolor, desde el miedo hasta la alegría" (Anil Seth, Being You: A New Science of Consciousness, 2020). Por lo tanto, la conciencia es la capacidad de sentir y experimentar.

El neurocientífico Michael Gazzaniga, profesor de psicología en la Universidad de California, en Santa Bárbara, muy reconocido por sus estudios sobre la conciencia, la define, de manera similar, como «la sensación subjetiva de un número de instintos y/o recuerdos desarrollándose en el tiempo dentro de un organismo».

Las «alucinaciones controladas»

Estos enfoques de la conciencia la vinculan estrechamente con nuestra condición de seres vivos, en contraposición a las posturas dualistas que la consideran una entidad separada del cuerpo o a las ideas panpsiquistas que la atribuyen a toda la materia, desde objetos inanimados hasta computadoras. La conciencia es, por tanto, un proceso biológico que nos hace ser algo más que objetos biológicos. Es lo que nos hace sentir vivos.

En el caso de los trabajos de Seth, uno de los conceptos más interesantes que propone es el de las «alucinaciones controladas». Se basa en que, aunque existe una realidad objetiva, no podemos experimentarla tal como es: nuestros cerebros construyen una interpretación de la realidad a partir de multitud de señales sensoriales, generando una percepción subjetiva. Nuestros cerebros han evolucionado para que estas alucinaciones controladas, estas interpretaciones que surgen del cerebro, sean bastante cercanas a la realidad, de modo que nos resulten útiles para sobrevivir.

Nuestra experiencia de la realidad es, en esencia, una construcción cerebral

El ejemplo más claro de esto es el color. En realidad, los colores no existen de manera objetiva. Son creaciones de nuestro cerebro basadas en cómo los objetos reflejan la luz. Estas interpretaciones son lo que experimentamos como color. Es una perspectiva que desafía la noción común de que percibimos el mundo «tal como es» y nos obliga a aceptar que nuestra experiencia de la realidad es, en esencia, una construcción cerebral.

Las investigaciones en neurociencia han demostrado que la percepción es el resultado de un equilibrio entre los datos sensoriales externos y las predicciones internas del cerebro. Este proceso se conoce como «codificación predictiva», donde el cerebro genera hipótesis constantes sobre el mundo y las ajusta según la información recibida. Cuando este mecanismo falla, pueden aparecer distorsiones perceptivas, como las ilusiones ópticas o los sueños o, incluso, trastornos como la esquizofrenia.

Los límites de la inteligencia artificial

Con los últimos avances en inteligencia artificial generativa, algunos han afirmado que las máquinas podrían alcanzar pronto una inteligencia artificial fuerte, es decir, que no simule tener estados mentales, sino que los tenga realmente y, por lo tanto, sea consciente. Personalmente, comparto con Anil Seth un alto grado de escepticismo sobre esta posibilidad. No hay ninguna evidencia científica que indique que una inteligencia artificial pueda llegar a ser consciente.

Es bien sabido que los humanos tenemos una tendencia natural a ser antropocéntricos, es decir, a ver el mundo desde nuestra propia perspectiva y a proyectar características humanas en otras entidades, en particular la capacidad de conciencia.

Un ejemplo claro son los grandes modelos de lenguaje de la inteligencia artificial. Están diseñados con la finalidad de producir una ilusión de inteligencia y personalidad. Su capacidad para generar lenguaje gramaticalmente correcto y su discurso persuasivo nos lleva a ver comprensión, intencionalidad y conciencia donde solo hay un proceso estadístico.

Las inteligencias artificiales muestran habilidades sin comprensión ni sensación de existencia propia

Es esta fuerte tendencia antropocéntrica la que lleva a muchos a cometer el error de sobrevalorar las

capacidades de los sistemas de inteligencia artificial o a creer que desarrollarán conciencia. Las inteligencias artificiales muestran habilidades sin comprensión, en el sentido de Daniel Dennett (From Bacteria to Bach and Back, 2018). Por muy sofisticadas e impresionantes que sean estas habilidades para resolver problemas, nunca tendrán experiencias subjetivas ni sensación de existencia propia por el hecho indiscutible de que no son seres vivos.

Procesadores de símbolos

En mi opinión, la Hipótesis del Sistema de Símbolos Físicos (SSF, por sus siglas en inglés) de Allen Newell y Herbert Simon (Computer Science as Empirical Inquiry: Symbols and Search, 1976) es falsa. Esta hipótesis postula que todo sistema capaz de procesar símbolos posee los medios necesarios y suficientes para ser inteligente en el sentido fuerte del término.

Aunque estrictamente la hipótesis SSF se formuló en 1975, ya estaba implícita en las ideas de los pioneros de la IA en los años 1950, e incluso en las ideas de Alan Turing en sus escritos sobre la posibilidad de la existencia de futuras máquinas inteligentes a finales de los años 40.

Conviene aclarar a qué se referían Newell y Simon cuando hablaban de Símbolos Físicos. Un SSF consiste en un conjunto de entidades llamadas símbolos que, mediante relaciones, pueden combinarse formando estructuras más grandes, como los átomos que se combinan formando moléculas, y que pueden ser transformados aplicando un conjunto de procesos. Estos procesos pueden introducir nuevos símbolos, crear y modificar relaciones entre ellos, almacenarlos, comparar si dos símbolos son iguales o diferentes, etc. Son símbolos físicos en tanto que tienen un sustrato electrónico (en el caso de los ordenadores) o biológico (en el caso de los seres humanos).

En definitiva, de acuerdo con esta hipótesis, la naturaleza del sustrato (circuitos electrónicos o redes neuronales) es indiferente a la hora de sustentar inteligencia, siempre y cuando este sustrato permita procesar símbolos. En el caso de los ordenadores, los procesadores de símbolos son lo que conocemos como programas.

La idea de la hipótesis SSF tiene sus raíces en la Teoría Computacional de la Mente que sostiene que la mente humana es un sistema de procesamiento de información y que la cognición y la conciencia, en conjunto, son una forma de computación. De hecho, John von Neumann, un matemático considerado uno de los padres de la informática, a mediados de los años 1940 ya comparaba el cerebro con el ordenador. En la década de 1960, Hilary Putman y Jerry Fodor defendían que la relación mente-cerebro era comparable con la relación software-hardware.

Sin embargo, esta analogía no es sostenible si tenemos en cuenta que la cognición no es como un programa de ordenador que se ejecuta en el cerebro, sino el resultado de millones de años de evolución. No se puede efectuar un cambio en la cognición sin modificar el cerebro, en cambio se puede cambiar el software sin modificar para nada el hardware. Se trata de una diferencia fundamental, aunque, obviamente, no es la única.

¿Conciencia sin biología?

Otra gran diferencia es que, contrariamente a un ordenador, la mayor parte de la información en el cerebro es construida en vez de almacenada. El cerebro, al revés que un ordenador, necesita almacenar poca información porque es capaz de focalizarse en lo esencial e inferir el resto con el fin de dar sentido a las cosas. Son estas construcciones las que probablemente dan lugar al sentido de identidad y a la consciencia.

Considerar cierta la hipótesis SSF implica creer que una inteligencia artificial igual o superior a la humana es posible. Contrariamente a esta creencia, estoy convencido de que la conciencia está profundamente arraigada

en la biología. Dicho de otra manera, el sustrato no sólo no es indiferente, sino que es definitivo. No podemos hablar de una auténtica comprensión del mundo y del lenguaje sin hablar de conciencia.

Una consecuencia importante es que las inteligencias artificiales, incluso las corpóreas, al no tener conciencia por el hecho de no ser seres vivos, no pueden comprender realmente ni el mundo ni el lenguaje, por mucho que nos parezca que los comprenden inducidos por nuestra tendencia a proyectar características humanas donde no las hay.

En conclusión, la conciencia sigue siendo un gran enigma. Su estudio es importante, no para construir inteligencias artificiales conscientes -ya que esto ni siquiera parece posible- sino porque, como dice Anil Seth, nos permitiría entender mejor nuestra propia naturaleza y mejorar la sociedad.

El simple hecho de reconocer que todos experimentamos el mundo de manera diferente podría ayudar a mejorar la empatía y la comunicación entre las personas

El simple hecho de reconocer que todos experimentamos el mundo de manera diferente podría ayudar a mejorar la empatía y la comunicación entre las personas. Entender que todo lo que percibimos es una construcción es una lección de humildad ante nuestras experiencias y creencias. A medida que avancemos en los estudios sobre la consciencia, es posible que descubramos no solo qué nos hace conscientes, sino también qué nos hace humanos.

Dennet, DC. (2018) From Bacteria to Bach and Back. PENGUIN. ISBN 9780141978048

Gazzaniga, M. (2018) The Consciousness Instinct: Unraveling the Mystery of How the Brain Makes the Mind, MacMillan Publishers. ISBN 9780374538156.

Hernández, S. "Anil Seth: «La realidad es una alucinación controlada" en CCCBLAB (2022). Disponible en https://lab.cccb.org/es/anil-seth-la-realidad-es-una-alucinacion-controlada/

Newell, A., Simon, HA. "Computer science as empirical inquiry: symbols and search" en Communications of the ACM (1976, Vol., Issue 3. Pages 113 – 126). Disponible en: https://doi.org/10.1145/360018.360022

Peñalver, J., González-García C. y Ruiz M. "La experiencia cambia la percepción: Codificación predictiva" en Ciencia Cognitiva (2023). Disponible en https://www.cienciacognitiva.org/?p=2296

Seth, A. (2020) Being You: A New Science of Consciousness. Faber and Faber. ISBN 9780571337729.

Zumalabe-Makirriain, JM. "El estudio neurológico de la conciencia: Una valoración crítica". en Anales de Psicología. (2016, Vol 32.1). Disponible en: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0212-97282016000100031