

Contemplándonos en el espejo de la inteligencia artificial

El desarrollo de la inteligencia artificial abre un nuevo camino para reflexionar y explorar posibles respuestas sobre algunos de los aspectos que más nos definen como seres humanos: nuestra inteligencia y nuestra consciencia.

Hace escasos años, términos como «atención», «sentimientos» o «alucinación», entre otros, se asociaban indefectiblemente a los seres humanos. Sin embargo, hoy en día, todos ellos se aplican de forma cotidiana también referidos a la inteligencia artificial (IA). Resulta significativo que uno de los artículos que más impacto ha tenido en su desarrollo reciente se titula simplemente "*Attention Is All You Need*"¹ (Atención es todo lo que necesitas). ¿Qué nos dice esto sobre nosotros mismos? ¿En qué medida la IA puede contribuir a explicar algunas de las cuestiones filosóficas sobre qué es lo que nos define como seres humanos?

La IA, como tal, constituye un campo amplio y complejo que abarca una gran cantidad de enfoques y técnicas: reconocimiento de patrones, visión por computadora, resolución de problemas matemáticos, robótica, juegos y un largo etcétera de facetas, no excluyentes. En todos los casos, para que podamos hablar de IA, la tecnología debe incorporar la capacidad de aprender, siendo este el factor clave que la define.

Los logros que se están consiguiendo con la IA son cada vez más espectaculares y acaban superando los hitos alcanzados por los propios seres humanos. Ello nos hace cuestionarnos en qué medida la forma en la que operan puede parecerse a la nuestra. El punto de partida resulta de por sí revelador, ya que las arquitecturas de redes neuronales, sobre las que se soportan en la actualidad, están inspiradas en las correspondientes a nuestra propia biología.

Los logros que se están consiguiendo con la IA nos hace cuestionarnos en qué medida la forma en la que operan puede parecerse a la nuestra

En este contexto, el hecho de que una red neuronal sea capaz de reconocer un objeto, por ejemplo, una silla, entre una multitud de imágenes, parece apuntar a que se ha formado una sólida noción sobre su identidad. Es decir, "sabe" lo que es una silla, concepto cuya descripción, desde el punto de vista filosófico, podría generar un interminable debate. La respuesta, en última instancia, se encuentra en una compleja función matemática con muchos parámetros ajustados. ¿Ocurre lo mismo en nuestro caso? ¿Son nuestros conceptos fórmulas matemáticas desarrolladas con la evolución y el aprendizaje? ¿Se aplica un mecanismo similar para otros tipos de procesos mentales? ¿En qué medida una red neuronal artificial (suponiendo que se construya con una arquitectura y algoritmos de entrenamiento lo suficientemente sofisticados) puede realizar los mismos procesos que nuestro cerebro?

¿En qué punto estamos?

Las inteligencias artificiales que se han venido desarrollando hasta ahora se han centrado en problemas específicos, siendo el reto futuro el de conseguir una IA de uso general que sea capaz de utilizarse en toda una variedad de problemas, de la misma forma que se aplica la inteligencia humana.

Quizá la rama de la IA que más se puede estar acercando a esta meta, hoy en día, es la del procesamiento del lenguaje natural, que se centra en algoritmos que permiten entender y generar lenguaje humano y que están detrás de aplicaciones como los asistentes virtuales o los modelos de lenguaje. Hay tres categorías de esfuerzos que están relacionados con su desarrollo: el modelado, el aprendizaje y el razonamiento, que está destinado a generar respuestas a preguntas que todavía no se han elaborado, mediante la manipulación del conocimiento actualmente existente junto con las técnicas de inferencia.

¿Podría considerarse que el dominio del lenguaje y su uso equivalente al que hace un ser humano pondría a las IA en pie de igualdad con nosotros? El lenguaje humano constituye una facultad extraordinaria que parece tener una componente genética, ya que los niños comienzan a hablar a parecidas edades, y una componente cultural, ya que los idiomas hablados dependen del entorno en el que cada uno se eduque. Si bien el lenguaje es una herramienta fundamental para el desarrollo de la inteligencia, parece que esta no estaría necesariamente ligada a su uso. Los bebés, sin manejar el lenguaje, son inteligentes y la inteligencia se despliega, en mayor o menor medida, a lo largo del reino animal. En línea con todo ello, el hecho de que una máquina sea capaz de manipular el lenguaje, y que incluso pueda pasar el Test de Turing, no tendría por qué significar necesariamente que haya adquirido las mismas capacidades que un ser humano.

¿Puede una IA llegar a ser consciente?

Esta discusión nos lleva a la que es, quizá, la cuestión más controvertida: la naturaleza de la consciencia y la posibilidad de que una inteligencia artificial pueda llegar a adquirirla. En este sentido, la transición de la inteligencia a la consciencia es una cuestión altamente debatida, sin que se hayan llegado todavía a alcanzar resultados concluyentes.

Algunos autores argumentan que la consciencia podría ser una propiedad emergente que surge cuando el nivel de inteligencia supera un determinado umbral de complejidad, lo que implicaría que sería explicable en función de esta, aunque no fuese reducible a los procesos inteligentes. Para ilustrar la brecha entre explicabilidad y reductibilidad cabe poner un ejemplo. Cinco puntos azules dispuestos de una forma determinada forman la T que es el logo de Telefónica. Este sería una propiedad emergente que surge de una agrupación específica de los puntos y, por lo tanto, es explicable en función de estos, pero no reducible a ellos.

Otra opción, sin entrar en consideraciones dualistas, es que la consciencia podría no tener mucho que ver con la inteligencia e, incluso, ser un fenómeno desligado de ella. La inteligencia se refiere a la capacidad de procesar información y realizar tareas de manera eficiente, mientras que la consciencia implica una experiencia subjetiva y una capacidad autorreferencial. Visto así la inteligencia vendría descrita con el verbo *hacer*, mientras que la consciencia con el verbo *ser*. Podría ocurrir, también, que la inteligencia fuese un prerrequisito para que surja la consciencia, pero que se requiera de algo adicional.

Para ahondar más en el problema de la consciencia, entraremos ahora en el terreno de David Chalmers que es, quizá, el filósofo que ha tenido una mayor influencia en las recientes discusiones sobre este tema y que propone dividir el problema en dos: el denominado «problema fácil» y el «problema difícil».

El «problema fácil» de la consciencia se referiría básicamente a las cuestiones relacionadas con sus correlatos neuronales. Es decir: ¿qué es lo que pasa en el cerebro cuando tenemos un determinado estado consciente?

El término «fácil» tendría que ver con el hecho de que es abordable mediante la investigación científica tradicional y la explicación de procesos neuronales.

El «problema difícil» se referiría a la cuestión fundamental de por qué surgen los estados mentales conscientes que experimentamos. En otras palabras: ¿por qué tenemos vida interior en lugar de simplemente procesar la información y actuar sin ser conscientes de ello? Por su propio planteamiento, este problema no es resoluble mediante la explicación de procesos físicos o neuronales, lo que justifica su calificación como “difícil”. Incluso si pudiéramos resolver el problema fácil y entender completamente el funcionamiento del cerebro, podríamos seguir sin entender por qué existen las experiencias conscientes.

¿Podría considerarse que el dominio del lenguaje y su uso equivalente al que hace un ser humano pondría a las IA en pie de igualdad con nosotros?

Para ilustrar estos conceptos entra en escena una criatura: el zombi filosófico. Este es un ser hipotético idéntico a un ser humano en términos de su comportamiento y funciones mentales, pero sin estar dotado de experiencias subjetivas. Es un «zombi» ya que puede realizar todas las acciones de la misma forma que lo haría un ser consciente, pero sin tener vida interior. Visto con una perspectiva evolutiva la pregunta que se plantea es: ¿por qué hemos llegado a ser seres humanos conscientes y no zombis filosóficos?

Las ideas de Chalmers han tenido una profunda influencia y han marcado los términos del debate estos últimos años, llevando a una fructífera reflexión sobre la naturaleza de la conciencia y la posible brecha entre los procesos físicos y la experiencia subjetiva.

Las conclusiones, sin embargo, siguen siendo esquivas. Si la conciencia es un fenómeno desligado del de la inteligencia o requiere de un algo adicional para manifestarse, por mucho que se desarrollen los sistemas de IA no llegarán a adquirirla. Ello, con independencia del énfasis con el que nos puedan manifestar lo contrario, si están entrenados para ello. Por otro lado, si la conciencia sí que puede surgir en los sistemas inteligentes, a partir de un umbral determinado de complejidad, sería posible que un día viésemos el nacimiento de un ser consciente artificial. Ello constituiría un hecho trascendental para la historia de la humanidad, cuyas consecuencias difícilmente podrían sobrestimarse.

Existiría, además, una implicación muy perturbadora, relacionada con la naturaleza de la realidad en la que habitamos. Si la conciencia fuese programable, y por lo tanto simulable, cualquier aspecto de la realidad podría llegar a serlo. Ello implicaría que sería factible crear realidades virtuales con seres conscientes como nosotros. Si esto fuese así, ¿podría ser que nuestra propia realidad fuera una construcción de este tipo y que estuviésemos habitando en un universo simulado?

Seth, A. (2023): *La creación del yo: Una nueva ciencia de la conciencia*. Madrid, Sexto Piso.

Chalmers, D. «Facing up to the problem of consciousness» en *Journal of consciousness studies*, 1995, vol. 2, no 3, p. 200-219.

Bostrom, N. «Are we living in a computer simulation?» en *The philosophical quarterly*, 2003, vol. 53, no 211, p. 243-255.