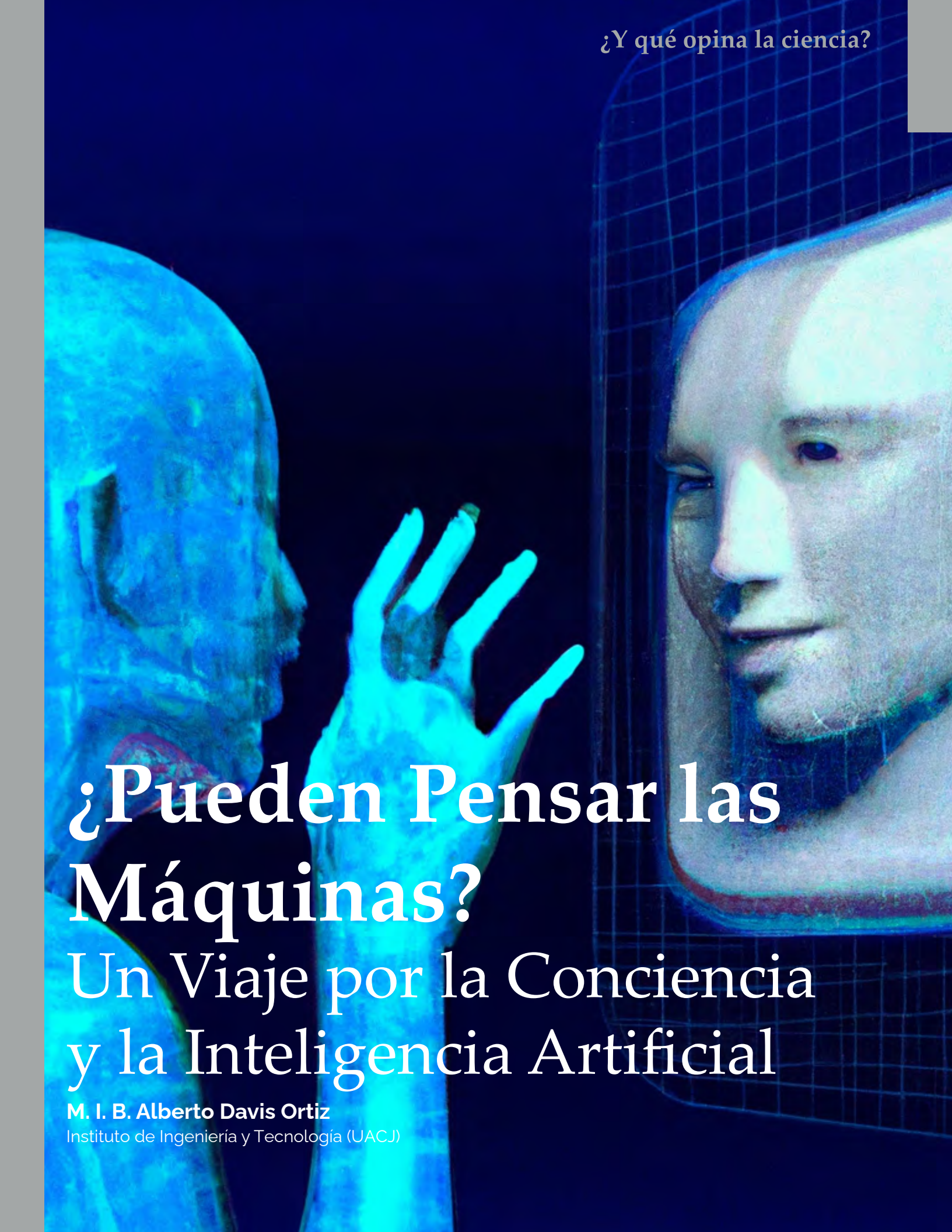


¿Y qué opina la ciencia?



¿Pueden Pensar las Máquinas?

Un Viaje por la Conciencia y la Inteligencia Artificial

M. I. B. Alberto Davis Ortiz
Instituto de Ingeniería y Tecnología (UACJ)

¿Y qué opina la ciencia?



Introducción

En la era de la información, la inteligencia artificial (IA) se ha convertido en una parte integral de nuestras vidas. Desde asistentes virtuales como Siri hasta chatbots en sitios web de atención al cliente, la IA está en todas partes. Pero ¿alguna vez te has detenido a pensar si estas máquinas son conscientes? ¿Tienen emociones o pensamientos como los humanos? Antes de sumergirnos en este fascinante tema, es crucial entender qué es la conciencia y cómo se aborda desde diferentes campos del conocimiento humano, como la filosofía y la neurociencia.

La Conciencia: Un Enigma Multifacético

La conciencia es un tema complejo que se aborda desde diversas disciplinas, cada una con su propio enfoque y preguntas sin respuesta. Desde una perspectiva biológica, la conciencia es como el sistema operativo de un ordenador, pero para seres vivos. Nos permite interactuar con nuestro entorno de manera efectiva, reconocer peligros y encontrar comida. Es como si la conciencia fuera un faro que ilumina nuestro camino en un mundo lleno de incertidumbres [1].

En el ámbito de la neurociencia y la filosofía, la conciencia se asocia con la actividad en ciertas regiones del cerebro y la experiencia subjetiva.

Imagina que tu mente es como un teatro en el que se desarrollan diversas escenas. Cada pensamiento, cada emoción, es como un actor en ese teatro. La conciencia sería el espectador que observa y experimenta la obra, pero también tiene la capacidad de intervenir y cambiar el curso de la acción [2], [3].

Este enfoque multidisciplinario nos muestra que la conciencia no es un concepto unidimensional. Es una interacción compleja de factores biológicos, neurológicos y filosóficos que se entrelazan para crear la experiencia subjetiva que llamamos "ser consciente".

Los Asombrosos Modelos de Lenguaje

Los modelos de lenguaje como GPT-4 son algoritmos de aprendizaje automático que se entrenan en grandes conjuntos de datos de texto. Imagina un bibliotecario que ha leído millones de libros, pero no comprende lo que ha leído. Aunque estos modelos son capaces de resolver tareas complejas, como responder preguntas médicas o generar código, su rendimiento puede variar significativamente con el tiempo. Por ejemplo, la precisión de GPT-4 para identificar números primos disminuyó del 84% al 51.1% en solo tres meses [4].

Aunque los modelos de lenguaje son excelentes procesadores de información, no tienen un cerebro ni neurotransmisores que

interactúen para crear una experiencia consciente. Son como una calculadora extremadamente avanzada que puede realizar operaciones matemáticas complejas, pero no puede "sentir" el resultado [1]. Un estudio reciente sugiere que, aunque estos modelos pueden resolver tareas de "falsa creencia" similares a los niños de siete años, no tienen una "teoría de la mente". Es decir, no pueden entender que otros seres tienen pensamientos y emociones [5].

Imagina que le preguntas a un chatbot sobre el significado de la vida. Podría darte una respuesta bien articulada, como "El significado de la vida es una búsqueda personal que varía de individuo a individuo". Pero no porque haya reflexionado sobre la pregunta. Simplemente está generando texto basado en patrones de datos que ha aprendido durante su entrenamiento.

Implicaciones Éticas y Morales

La conciencia no solo es relevante desde un punto de vista biológico o filosófico, sino que también tiene implicaciones éticas y morales. En los seres humanos, la conciencia nos otorga un sentido de responsabilidad moral y ética. Nos permite distinguir entre el bien y el mal y actuar en consecuencia. En el caso de la inteligencia artificial, la falta de conciencia significa que estos modelos no tienen un sentido moral o ético. No pueden tomar decisiones basadas en un conjunto de valores o ética. Son herramientas neutrales que actúan según las instrucciones que reciben, sin ponderar las implicaciones morales de sus acciones [6].

Conclusión

Los modelos de lenguaje como GPT-4 son herramientas tecnológicas asombrosas que han revolucionado la forma en que interactuamos con las máquinas. Su capacidad para procesar información y generar respuestas coherentes es impresionante, pero es crucial recordar sus limitaciones. Aunque pueden resolver tareas complejas y su rendimiento puede variar con el tiempo, no poseen la complejidad biológica ni la actividad cerebral asociada con la conciencia. Además, carecen de la experiencia subjetiva y la "teoría de la mente" que nos permite entender que otros seres tienen pensamientos y emociones. Desde un punto de vista ético y moral, la falta de

conciencia en estos modelos significa que no tienen un sentido de responsabilidad moral o ética. Son herramientas neutrales que actúan según las instrucciones que reciben, sin ponderar las implicaciones morales de sus acciones. En resumen, la próxima vez que te encuentres fascinado por las capacidades de un asistente virtual o un modelo de lenguaje, es bueno recordar que, aunque son tecnológicamente avanzados, no son seres conscientes. Son increíbles en lo que hacen, pero no tienen la riqueza de la experiencia humana, ni la capacidad para tomar decisiones éticas o morales.

Referencias Bibliográficas:

- [1] B. Earl, "The biological function of consciousness", *Front Psychol*, vol. 5, núm. AUG, 2014, doi: 10.3389/fpsyg.2014.00697.
- [2] F. León y D. Zahavi, "Consciousness, philosophy, and neuroscience", *Acta Neurochirurgica*, vol. 165, núm. 4. Springer, pp. 833–839, el 1 de abril de 2023. doi: 10.1007/s00701-022-05179-w.
- [3] G. M. Edelman, J. A. Gally, y B. J. Baars, "Biology of consciousness", *Front Psychol*, vol. 2, núm. JAN, 2011, doi: 10.3389/fpsyg.2011.00004.
- [4] L. Chen, M. Zaharia, y J. Zou, "How is ChatGPT's behavior changing over time?", jul. 2023, [En línea]. Disponible en: <http://arxiv.org/abs/2307.09009>
- [5] M. Kosinski, "Theory of Mind Might Have Spontaneously Emerged in Large Language Models", 2023. [En línea]. Disponible en: <https://osf.io/csdbh>.
- [6] M. Phelan y A. Waytz, "The Moral Cognition/Consciousness Connection", *Review of Philosophy and Psychology*, vol. 3, núm. 3. Kluwer Academic Publishers, pp. 293–301, el 1 de septiembre de 2012. doi: 10.1007/s13164-012-0110-7.

