



GABRIELA
ARRIAGADA



ABEL
WAJNERMAN

Los errores y las limitaciones detrás de los algoritmos de IA: una oportunidad para revalorar lo humano.



GABRIELA ARRIAGADA BRUNEAU

Profesora Asistente, Instituto de Éticas Aplicadas
Instituto de Ingeniería Matemática y Computacional
Pontificia Universidad Católica de Chile

ABEL WAJNERMAN PAZ

Profesor Asistente,
Instituto de Éticas Aplicadas
Pontificia Universidad Católica de Chile

Las inteligencias artificiales (IAs) han irrumpido en nuestras vidas con una fuerza arrolladora, transformando desde la medicina hasta la educación, pasando por el entretenimiento, la ciencia y la industria. Con cada nueva innovación, se nos presenta una visión de un futuro más eficiente, más rápido y, en muchos casos, más preciso que el presente, generando nuevas dependencias o confianzas en la tecnología, como el fenómeno de la apreciación algorítmica¹ que implica que los seres humanos a menudo demuestran una mayor confianza en los consejos provenientes de algoritmos en comparación con los consejos no algorítmicos. Desde una perspectiva ética, esta confianza puede ser problemática porque conlleva una creciente delegación de actividades significativas a la IA, potencialmente limitando o degradando experiencias humanas fundamentales.

Un peligro clave de delegar de esta manera en la IA reside en los tipos de errores que cometen los algoritmos de aprendizaje profundo, los que ponen de relieve la profundidad de nuestra mente para comprender y abordar desafíos humanos. Mientras que los humanos tienen una capacidad inherente para interpretar contextos complejos, los algoritmos de IA, a pesar de su capacidad para procesar y analizar grandes volúmenes de datos e identificar patrones que son invisibles a la mente humana, a menudo fallan en situaciones que un ser humano consideraría triviales². Tomemos, por

ejemplo, un experimento en el cual una IA, entrenada para reconocer imágenes, no logra identificar un pingüino cuando la imagen está invertida. Para un humano, la imagen invertida de un pingüino sigue siendo reconocible como tal; nuestro cerebro automáticamente corrige la orientación y recupera el significado esencial de la figura. Pero para la IA la inversión de la imagen puede resultar en un error catastrófico, revelando una falla en la comprensión subyacente.



Ilustración 1:
"En 2024 aún era difícil para la IA reconocer el pingüino rey aquí".

1 Jennifer M. Logg, Julia A. Minson, and Don A. Moore, 'Algorithm Appreciation: People Prefer Algorithmic to Human Judgment', *Organizational Behavior and Human Decision Processes* 151, no. C (2019): 90-103.

2 J. E. (Hans). Korteling et al., 'Human - versus Artificial Intelligence', *Frontiers in Artificial Intelligence* 4 (25 March 2021): 622364, <https://doi.org/10.3389/frai.2021.622364>.

Este tipo de error epistemológico es sintomático de una diferencia fundamental entre la cognición humana y los algoritmos de IA. Mientras que los humanos operan en un mundo de significados y contextos, donde los errores suelen ser el resultado de malinterpretaciones o falta de información, los errores de la IA son más estructurales y técnicos. Los algoritmos de IA no “entienden” el mundo; más bien, operan basados en correlaciones y patrones estadísticos extraídos de los datos que se les proporcionan. Esto crea un problema de confianza epistémica: ¿cómo podemos confiar plenamente en sistemas que, aunque son capaces de superar nuestras capacidades en ciertas tareas, fallan de manera tan básica en otras?

Otro tipo de limitación clave se refleja en contextos donde la empatía y la conexión emocional son importantes, como la educación, la salud y las relaciones personales. Las IAs que imitan interacciones humanas en estos contextos para, por ejemplo, cuidar y apoyar a pacientes médicos tienen la capacidad de detectar o reconocer estados emocionales a partir de expresiones verbales o faciales. Estos dispositivos pueden ser de gran ayuda para apoyar a profesionales de la salud altamente demandados. Sin embargo, carecen de la empatía emocional que nos permite experimentar en carne propia las emociones sentidas por otras personas. Esta capacidad humana para resonar emocionalmente con otros es necesaria para tener una comprensión profunda y contextualizada de sus necesidades y, de esta manera, perseguir competentemente nuestros deberes e ideales de cuidado³.

De este modo, se ve que las limitaciones de la IA van más allá de cuestiones asociadas con la mera eficiencia técnica. Los errores y limitaciones de la IA nos ofrecen una oportunidad para reflexionar sobre lo que significa ser humano, donde podemos revalorar la importancia de las subjetividades y experiencias en nuestro quehacer. Las capacidades humanas de contextualizar la información y procesarla emocionalmente constituyen un valor que no puede ser replicado por la IA, al menos no en su forma actual. Aquí es donde surge la oportunidad de revalorar lo humano. En lugar de ver los errores de la IA como meros obstáculos a superar, deberíamos considerarlos como recordatorios de las limitaciones de la tecnología y la importancia de mantener un enfoque centrado en el ser humano.

Además, estos errores nos instan a cuestionar la dirección en la que estamos llevando el desarrollo tecnológico. La obsesión por la eficiencia y la precisión algorítmica corre el riesgo de deshumanizar procesos que, por su naturaleza, requieren una comprensión profunda del contexto y el significado. Es crucial que, en nuestro afán por avanzar tecnológicamente, no perdamos de vista la necesidad de mantener la humanidad en el centro de nuestras innovaciones. Preguntas clave debiesen apuntar a cómo desplegar sistemas de IA de manera efectiva para complementar y compensar las limitaciones inherentes a la cognición humana (y viceversa), entendiendo estas limitaciones como parte fundamental e inherente del desarrollo de las IAs. Es ese balance entre antropocentrismo y antropomorfismo el que debemos cuidar. [®]

ple Obstacles for Empathic AI: Why We Can't Replace Human Empathy in Healthcare', *AI & SOCIETY* 37, no. 4 (1 December 2022): 1353-59, <https://doi.org/10.1007/s00146-021-01230-z>.

3 Carlos Montemayor, Jodi Halpern, and Abrol Fairweather, 'In Princi-