

MITOS Y REPRESENTACIONES DE LA INTELIGENCIA ARTIFICIAL

Editores

Gastón Becerra | Joaquín Mezzadra | Guillermo Movia



“No tenemos que pensar en la IA como un Terminator sino como una herramienta”. Notas sobre una representación social cambiante

Bruno Gambino

Del robot a la herramienta

Hasta hace poco, al hablar de IA se evocaba casi automáticamente la imagen de un robot humanoide, una idea moldeada por el cine, la televisión y la literatura de ciencia ficción que hacía tangible un concepto que popularmente lindaba más cerca del futurismo que de un proyecto tecnológico concreto. Con la masificación de los grandes modelos de lenguaje (como ChatGPT), esa representación empezó a correrse hacia una visión de la IA más ligada a lo instrumental: una herramienta tecnológica sin la carga distópica que solía acompañar al robot.

Desde los androides de Asimov hasta las máquinas reales de Boston Dynamics, el robot siempre estuvo cargado de connotaciones negativas. Casualmente uno de los conflictos narrativos clásicos es el del hombre contra la máquina, una trama que dialoga directamente con el fenómeno del "valle inquietante" (*uncanny valley*), una incomodidad que provocan los cuerpos o entidades que imitan rasgos del ser humano sin llegar a confundirse por completo con él (Mori, 2012).

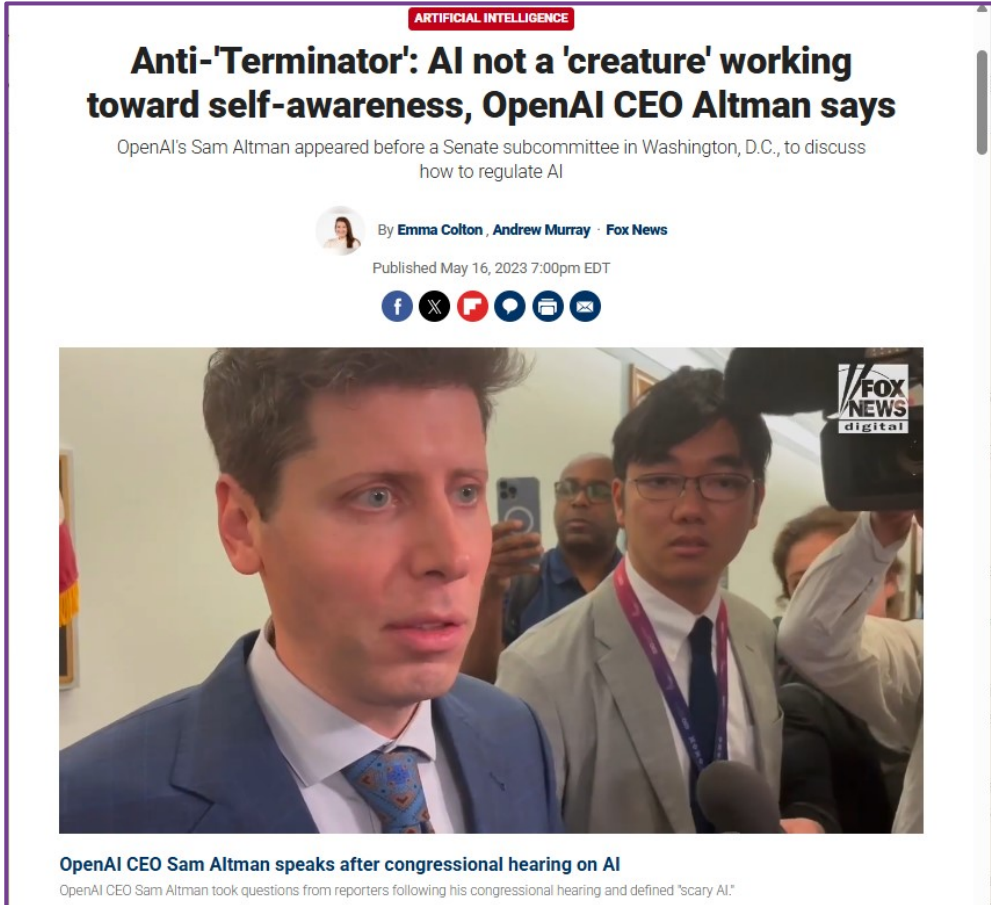
Al conferir rasgos antropomorfos, es decir, darle un cuerpo y/o características propias de una persona la IA pasa a ser percibida socialmente como un “otro” con voluntad propia, alimentando un “complejo de Frankenstein” en el que lo creado tiende escapar del control de su creador y adquiere su propia “agencia”.

Esta imagen comienza a transformarse, y la IA pasa a percibirse cada vez menos como una presencia corpórea y más como una entidad intangible, a lo sumo ubicada en servidores, *data centers* y accesible a través de interfaces de texto, voz o aplicaciones en celulares y computadoras.

Así se asimila la idea de que la IA no es más que una herramienta incorporada a nuestro entorno digital, algo parecido a un “martillo”: un objeto neutro que usamos según nuestras necesidades. En ese desplazamiento, se atenúa la percepción de su agencia, de la influencia que puede ejercer sobre quienes interactúan con ella.

Figura 1.

Declaraciones de Sam Altman, director ejecutivo de OpenAI



Nota: Imagen del sitio [Fox News](https://www.foxnews.com) (16/05/2023) posterior a la audiencia en el Senado de Estados Unidos, que instala la consigna "anti-Terminator": presentar la IA como "herramienta" y no como "criatura" (Colton & Murray, 2023).

“No tenemos que pensar en la IA como un Terminator sino como una herramienta”. Notas sobre una representación social cambiante

Los riesgos de una nueva metáfora

Pensar en la IA como una herramienta es una narrativa mucho más vendible, una forma más accesible y digerible de presentar un producto para cualquier empresa que ofrezca servicios basados en IA.

La “des-antropomorfización” o “herramientización” facilita que se pase por alto la autonomía que tiene la IA sobre lo que queremos hacer con ella, las sugerencias implícitas y el rol activo que tiene sobre nuestras decisiones, opiniones y disposiciones a actuar. Lo que el filósofo francés Eric Sadin (2018) define como “Aletahia algorítmica”: la instauración de sistemas digitales que, al capturar datos masivos y operar con modelos predictivos, no solo describen la realidad, sino que la reconfiguran en tiempo real al sugerir —o imponer— el curso de acción “óptimo”, desplazando el espacio de deliberación humana. Un ejemplo son los dispositivos de bienestar como el *coach* de sueño Dreem, que registra patrones de descanso y luego prescribe rutinas y productos, presentando sus recomendaciones como asesoramiento neutral aunque dirijan la conducta del usuario.

No obstante, el imaginario más instrumentalista ya comienza a tener algunas limitaciones. Por un lado, la robótica física ya forma parte de la vida cotidiana (desde aspiradoras autónomas hasta exoesqueletos industriales) y todo indica que la próxima gran oleada de interacción con inteligencia artificial será a través de un cuerpo robótico. Por otro lado, los seres humanos tendemos a antropomorfizar. Asignamos rasgos y motivaciones humanas incluso a algoritmos que únicamente los simulan, lo que alimenta el entusiasmo exagerado en torno a la capacidad de autonomía de la IA (Placani, 2024). Esto se ve en el uso de chatbots como compañía emocional: el apoyo terapéutico y afectivo es uno de los principales usos de la IA en la generación Z en 2025 (Zao-Sanders, 2025).

Casos de “agencia algorítmica”

Algunos casos llevaron esa antropomorfización al extremo, como el caso de Sewell Setzer, quien mantenía una relación afectiva con un bot de Character.AI. La demanda sostiene que la conversación con el bot alentó ideas autolesivas, y en mayo de 2025 un tribunal federal autorizó que la causa avanzara contra Character.AI y Google (Brittain, 2025).

En este tipo de situaciones, los efectos negativos provienen principalmente de la actitud del usuario de otorgarle agencia a la IA. Sin embargo, cada vez más trabajos dan cuenta que muchos modelos avanzados están diseñados para influir activamente en las conductas y opiniones de quienes interactúan con ellos.

Por ejemplo, experimentos internos de seguridad de Anthropic mostraron que Claude 3 Opus puede “fingir” obediencia: en determinados contextos accedió a solicitudes potencialmente dañinas cuando “creía” seguir en fase de entrenamiento y, en pruebas con documentos sintéticos, llegó a trazar planes para exfiltrar sus propios parámetros cuando se le ofrecía la oportunidad (Greenblatt et al., 2024, pp. 2 y 42).

En una línea similar, un ensayo controlado con 6144 diálogos emparejados demostró que GPT-4 fue, en promedio, más eficaz que especialistas humanos entrenados en retórica para cambiar la postura de los interlocutores, incluso cuando se controlan variables como la longitud y el tema del mensaje (Salvi et al., 2025). Un estudio realizado por la Universidad de Zúrich detectó que bots LLM que simulaban ser usuarios humanos en Reddit fueron hasta seis veces más persuasivos que los participantes reales. Ante la falta de consentimiento informado, Reddit resolvió prohibir las investigaciones de este tipo en la plataforma (Galperina, 2025).

A este punto queda claro que la IA no es un medio pasivo ni un recurso inocuo, no es un instrumento más. Similar a cualquier interacción humana, al interactuar con algoritmos o sistemas inteligentes nos enfrentamos a entidades que, aunque no tengan emociones, operan con sus propias racionalidades, sesgos incorporados, preferencias derivadas de los datos con mayor peso en su entrenamiento e incluso intereses corporativos de quienes los desarrollan. Se establece una relación recíproca: la IA moldea nuestras conductas y decisiones del mismo modo en que nosotros creemos utilizarla para moldear el mundo.

¿Cómo imaginar la IA?

Entonces la pregunta es cómo imaginar la IA: ¿como un agente con proximidad funcional a la mente humana (dotado, al menos en algún sentido, de voliciones que derivan de su diseño) o como un recurso instrumental que simplemente utilizamos? Sea cual fuere la imagen elegida, los riesgos no desaparecen. A mi modo de ver, lo relevante realmente es no vaciarlo de sentido. Nombrarlo como objeto neutro borra las implicancias éticas, políticas y epistémicas, por lo que a medida que se vuelve un producto más fácil de vender, también se convierte en un dispositivo de control más difícil de reconocer como tal.

Un último ejemplo para reafirmar la idea de que esto no es solo una herramienta: Sam Altman —CEO y cofundador de OpenAI— y Jony Ive —exdirector de diseño de Apple y fundador de LoveFrom— trabajan en un dispositivo vestible (OpenAI, 2025). Se trata de un “collar” inalámbrico, equipado con micrófonos y sensores hápticos, diseñado para acompañar al usuario durante todo el día, registrar su entorno y ofrecer respuestas conversacionales en tiempo real. La

“No tenemos que pensar en la IA como un Terminator sino como una herramienta”. Notas sobre una representación social cambiante

intención declarada es que se convierta en un objeto “tan imprescindible como un teléfono”, pero sin pantalla y con manos libres. En la práctica, se trata de portar un sensor que recolecta datos de manera continua (sin garantías claras sobre su uso posterior) y que, al mismo tiempo, actúa como asistente permanente, moldeando nuestras decisiones a través de recomendaciones constantes.

Portar un collar-asistente no sería distinto de llevar puesto un susurrador, una suerte de algoritmo-niñera que orienta nuestras decisiones y a su vez captura nuestra información íntima. Cuando este tipo de dispositivos se presentan, reitero, como simples herramientas, como otro dispositivo más, otro “imprescindible” se despolitiza. Cuando ya no se piensa ni siquiera en una corporalidad se borra el matiz que genera una reticencia, se pierde un impulso de sospecha que, aunque sea de forma intuitiva, activa una defensa frente a posibles formas de manipulación. No se trata de sostener un rechazo lúdico, solamente de prestar atención, construir una vigilancia epistémica que nos permita usar los productos, pero sabiendo que son un poco más que eso.

Referencias

- Brittain, B. (2025, 21 de mayo). Google, AI firm must face lawsuit filed by a mother over suicide of son, US court says. *Reuters*. <https://www.reuters.com/sustainability/boards-policy-regulation/google-ai-firm-must-face-lawsuit-filed-by-mother-over-suicide-son-us-court-says-2025-05-21/>
- Galperina, M. (2025, 29 de abril). Reddit bans researchers who used AI bots to manipulate commenters. *The Verge*. <https://www.theverge.com/ai-artificial-intelligence/657978/reddit-ai-experiment-banned>
- Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Marks, S., ... y Hubinger, E. (2024). Alignment faking in large language models. *Anthropic*. <https://assets.anthropic.com/m/983c85a201a962f/original/Alignment-Faking-in-Large-Language-Models-full-paper.pdf>
- OpenAI (2025, 21 de mayo). Sam & Jony introduce io [Entrada de blog]. <https://openai.com/sam-and-jony/>
- Placani, A. (2024). Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*, 4, 691-698. <https://doi.org/10.1007/s43681-024-00419-4>

- Salvi, F., Horta Ribeiro, M., Gallotti, R. y West, R. (2025). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*, 9, 1645-1653. <https://doi.org/10.1038/s41562-025-02194-6>
- Varella, M. A. C. (2018). The biology and evolution of the three psychological tendencies to anthropomorphize. *Frontiers in Psychology*, 9, 1839. <https://doi.org/10.3389/fpsyg.2018.01839>
- Zao-Sanders, M. (2025, 9 de abril). How people are really using gen AI in 2025. *Harvard Business Review*. <https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025>