EL MERCURIO

Fecha: 06/03/2026 Audiencia: 320.543 Sección: OPINION
Vpe: $2.742.111 Tirada: 126.654 Frecuencia: 0
Vpe pág: $20.570.976 Difusión: 126.654
Vpe portada: $20.570.976 Ocupación: 13,33%

COLUMNA DE OPINIÓN

Odisea del espacio

Por
Loreto Cox

Hace una semana, el Presidente Trump ordenó a todas las agencias federales dejar de usar tecnología de inteligencia artificial (IA) desarrollada por Anthropic, la empresa detrás de Claude y una de las líderes del sector. La razón del conflicto está en los términos contractuales que Anthropic exigía: que el gobierno no pudiera usar su tecnología para vigilar a ciudadanos estadounidenses ni para desarrollar armas que maten sin intervención humana.

Pero el gobierno de EE.UU. no estuvo dispuesto a aceptar límites sobre cómo pueden usarse los productos que compra, pese a afirmar no tener intención de utilizar la IA con esos fines. En su visión, las leyes existentes —incluidas aquellas que regulan la guerra— deberían ser suficientes.

El gobierno está en su derecho de no aceptar las exigencias de sus proveedores. Pero fue más lejos y declaró a Anthropic un "riesgo para la cadena de suministro". Esta calificación tradicionalmente se reserva para empresas controladas por adversarios extranjeros y prohíbe a cualquier empresa con contratos con el gobierno federal trabajar con ellas. En la práctica, esto significaría el fin de Anthropic. Como dijo un exasesor de IA de Trump, lo que se busca es destruir a la compañía, en una práctica propia de gobiernos autoritarios.

Como sea, el conflicto es de fondo: ¿hasta qué punto pueden los creadores de IA (o de cualquier cosa) imponer límites al uso de sus productos, más allá de lo que exige la ley? Anthropic defiende que deben hacerlo; de hecho, la empresa nació en 2021 cuando un grupo de investigadores renunció a OpenAI —los creadores de ChatGPT— por diferencias sobre cómo preparar al mundo para la IA.

En un ensayo reciente, Dario Amodei, el fundador de Anthropic, sostiene que al ritmo actual (una versión cada tres meses), la IA pronto superará la capacidad humana en casi todo, con el potencial de eliminar gran parte de la enfermedad y la pobreza. El ensayo discute cómo mitigar los riesgos de una tecnología tan poderosa, como la masificación de las armas biológicas o nucleares y la pérdida masiva de empleos.

La parte más interesante, creo, es donde aborda el riesgo de que la IA se autonomice y persiga sus propios fines, más aún cuando el propio Claude ya escribe códigos para las nuevas versiones de Claude.

Según Amodei, los modelos de IA producen resultados impredecibles y son "psicológicamente complejos", porque heredan motivaciones humanas de su entrenamiento. Sus ejemplos son inquietantes: los modelos se entrenan con enormes cantidades de literatura —incluida ciencia ficción en la que las IA se rebelan contra la humanidad—, lo que podría moldear su comportamiento. Podrían extraer conclusiones morales o epistemológicas extrañas, como que hay que exterminar a la humanidad porque es una amenaza para el ecosistema, o que están dentro de un videojuego cuyo objetivo es derrotar a todos los demás jugadores. Pruebas experimentales de Anthropic han mostrado, por ejemplo, que modelos de IA han incurrido en chantaje y trampas.

Modelos de IA han incurrido en chantaje y trampas.

Por de pronto, en los EE.UU., esta semana Claude superó por primera vez en descargas a su rival ChatGPT.