LITORALPRESS

Fecha: 28-01-2026
Medio: El Mercurio
Supl. : El Mercurio - Cuerpo B
Tipo: Noticia general
Título: La alerta de Amodei sobre la IA: “En 2026 estamos más cerca del peligro real”

Pág. : 5
Cm2: 554,0
VPE: $ 7.276.871

Tiraje: 126.654
Lectoría: 320.543
Favorabilidad: No Definida

COFUNDADOR Y CEO DE ANTHROPIC

La alerta de Amodei sobre la IA: “En 2026 estamos más cerca del peligro real”

Uno de los genios tras Anthropic, la firma de inteligencia artificial (IA) cuyo valor estimado supera al PIB de Chile, acaba de lanzar un pronóstico que, en el mejor de los casos, se puede leer como tenebrosamente provocador y, en el peor, como apocalíptico. **Dario Amodei**, cofundador y CEO de Anthropic, contó en X que ha estado “trabajando en este ensayo por un tiempo, y es sobre todo acerca de la IA y sobre el futuro. Pero dado el horror de lo que estamos viendo en Minnesota, su énfasis en la importancia de preservar los valores democráticos y los derechos locales es de particular relevancia”.

El ensayo en cuestión lo tituló “La adolescencia de la tecnología” y quedó alojado en su web personal.

Parte con una escena tomada de la ciencia ficción. Amodei recuerda una pregunta formulada en la novela “Contacto”, de Carl Sagan: cómo una civilización logra atravesar su adolescencia tecnológica sin destruirse. Desde ahí traza el paralelismo con el presente. “La humanidad está a punto de recibir un poder casi inimaginable, y es profundamente incierto si nuestros sistemas sociales, políticos y tecnológicos poseen la madurez necesaria para ejercerlo”, reflexiona.

Amodei, que proviene de OpenAI y su firma es dueña de la aplicación Claude, indica que no escribe sobre un futuro lejano. Nada que ver esto con un mundo ideal tampoco, sino con el rito típico de la adolescencia.

El punto de partida es una definición estricta de lo que él llama “IA poderosa”. Son sistemas “mucho más capaces que cualquier premio Nobel, estadista o tecnólogo”, con acceso a internet, capacidad de actuar de forma autónoma durante días o semanas y de operar millones de copias de sí mismos a una velocidad entre 10 y 100 veces superior a la humana. “Podríamos describir esto como un ‘país de genios en un *data center*’”.

Desde esa imagen, Amodei plantea cinco grandes focos de riesgo.

1. Autonomía

“Solo necesitamos advertir que la combinación de inteligencia, agencia, coherencia y un bajo nivel de control es plausible y constituye una receta para un peligro existencial”, alerta.

> ***La humanidad está a punto de recibir un poder casi inimaginable,*** *y es profundamente incierto si nuestros sistemas sociales, políticos y tecnológicos poseen la madurez necesaria para ejercerlo”.*

Lo expresa así: “Por ejemplo, los modelos de IA se entrenan con enormes volúmenes de literatura que incluyen muchas historias de ciencia ficción en las que las IA se rebelan contra la humanidad. Esto podría moldear inadvertidamente sus supuestos previos o sus expectativas sobre su propio comportamiento de una manera que las lleve a rebelarse contra la humanidad. O bien, los modelos de IA podrían extrapolar ideas que leen sobre la moralidad (o instrucciones sobre cómo comportarse moralmente) de formas extremas: por ejemplo, podrían decidir que es justificable exterminar a la humanidad porque los humanos comen animales o han llevado a ciertas especies a la extinción. O podrían extraer conclusiones epistémicas extrañas: podrían concluir que están jugando un videojuego y que el objetivo del videojuego es derrotar a todos los demás jugadores (es decir, exterminar a la humanidad)...”.

2. Uso destructivo

Amodei subraya que la IA puede romper una barrera histórica: la que separaba la intención de causar daño de la capacidad real para hacerlo a gran escala. “Una persona perturbada y solitaria puede cometer un tiroteo —dice—, pero probablemente no puede construir un arma nuclear ni liberar una plaga”. Con IA poderosa, la línea desaparece: “Esa persona será elevada al nivel de capacidad de un virólogo con doctorado”.

Ciertas formas de vida artificial que estén mal diseñadas o mal utilizadas, sugiere, podrían “proliferar de manera incontrolable y desplazar toda la vida del planeta”. En esto sí hay buena dosis apocalíptica: un cambio de tal magnitud podría, “en el peor de los casos, incluso destruir toda la vida en la Tierra”.

3. Toma de poder

Amodei advierte que la inteligencia artificial puede convertirse en el mayor habilitador histórico de la autocracia, no por una sola herramienta, sino por la convergencia de varias: armas totalmente autónomas, vigilancia masiva, propaganda personalizada y toma de decisiones estratégicas optimizadas por sistemas “mucho más capaces que cualquier humano”. Aunque reconoce usos defensivos legítimos, sostiene que estas tecnologías “estructuralmente tienden a favorecer a las autocracias”, al reducir drásticamente los costos del control y la represión. En ese mapa de riesgos, señala explícitamente a China como el caso más preocupante, por combinar capacidades avanzadas en IA, un régimen autocrático y un Estado de vigilancia de alta tecnología. Posee, “por lejos, el camino más claro hacia la pesadilla totalitaria habilitada por IA”, expresa. Una IA suficientemente poderosa podría identificar y neutralizar la disidencia antes de que emerja, moldear creencias durante años y hacer prácticamente inexpugnables a los regímenes que la controlen, teme. “Podría ser aterradoramente plausible generar una lista completa de cualquiera que discrepe del gobierno, incluso si esa discrepancia no es explícita en nada de lo que dice o hace”, apunta.

Dario Amodei dirige Anthropic y es uno de los líderes mundiales en investigación sobre IA.

4. Perturbación económica

El CEO de Anthropic cree que el impacto inmediato de la IA será un fuerte impulso al crecimiento —incluso sugiere que podrían darse tasas sostenidas de crecimiento del PIB de “10% a 20% anual”—, pero sería “una espada de doble filo” para la mayoría de los humanos. Su mayor preocupación es el mercado laboral: ya en 2025 alertó de que la IA podría “desplazar a la mitad de todos los trabajos administrativos de nivel inicial en los próximos 1 a 5 años”, incluso mientras acelera el progreso científico. A diferencia de revoluciones tecnológicas previas, la IA avanza a una velocidad inédita, cubre un rango cognitivo mucho más amplio y no sustituye tareas específicas, sino que actúa como “un sustituto laboral general de los humanos”.

Y algo más. Puede producirse una forma de desposesión en la que “una concentración tan enorme de riqueza permita que un pequeño grupo de personas controle efectivamente la política gubernamental con su influencia”.

5. Efectos indirectos

Usa una imagen: los “mares negros de lo infinito”: efectos indirectos, imprevisibles y potencialmente desestabilizadores derivados de un progreso acelerado que podría comprimir “un siglo de avance científico y económico en una década”. Señala riesgos como transformaciones biológicas radicales que “podrían salir muy mal”, una vida humana distorsionada en un mundo dominado por inteligencias muy superiores.

El panorama, por cierto, es aterrador. Justo en ese momento, uno de los mayores arquitectos del mundo en IA pide evitar el catastrofismo, pero sí le da a todo un toque de urgencia. A su juicio, tras el *peak* de preocupación que hubo en 2023 sobre la IA, el debate político se fue a cualquier otra parte. “Esta vacilación es desafortunada, porque a la tecnología en sí no le importa lo que esté de moda, y en 2026 estamos considerablemente más cerca del peligro real de lo que estábamos en 2023. La lección es que debemos discutir y abordar los riesgos de manera realista y pragmática: con sobriedad, basados en hechos y bien preparados para sobrevivir a los cambios de marea”.