La IA puede salvarnos… de la Estupidez Natural

"We are being lied to… We are told to be pessimistic", advierte Marc Andreessen (2023) en su *Manifiesto tecno-optimista*. Prometeo reaparece como Frankenstein, Oppenheimer, Terminator o HAL: cambia el monstruo, pero persiste la sospecha de que seremos castigados por transformar la naturaleza.

Jacob Coxon, Evan Hubinger y Dario Amodei advierten que sistemas capaces de mejorarse podrían escapar del control humano y exterminarnos (Reuters, 2026a). La posibilidad merece estudio; convertirla en certidumbre política es otra cosa. Amodei (2026) propone ralentizar las capacidades de frontera para que la seguridad pueda alcanzarlas.

Sin embargo, como señala Londoño (2026), varios episodios de supuesta pérdida de control, invocados para justificar la alarma ante la IA, reflejan errores humanos de diseño experimental. Cuanto más se intensifica esa alarma, más se soslaya el riesgo opuesto: frenar la tecnología que permite descubrir vulnerabilidades puede inmovilizar a los defensores mientras sus adversarios continúan aprendiendo.

La alarma, además, ha adquirido utilidad electoral. Un análisis de 1,200 sitios de campaña encontró referencias a la IA o a centros de datos en casi 40% de las contiendas estadounidenses; los candidatos demócratas los abordaban más del doble que los republicanos (Ovide et al., 2026).

Barack Obama ha instado al Partido Demócrata a convertir la IA en una prioridad electoral (Reuters, 2026b). Sin duda, nada de esto prueba una conspiración, pero sí permite advertir oportunismo electoral: cuando el miedo promete votos, la prudencia puede transformarse en una subasta de prohibiciones.

Pero eso no es todo: la evidencia publicada por la propia *Anthropic* plantea un dilema aún más incómodo. Una codificación conservadora de los 42 estudios de caso incluidos en su informe de septiembre de 2026 sobre el uso malicioso de la IA arroja la siguiente distribución (Anthropic, 2026):

Tabla 1

Distribución jurisdiccional de los casos publicados por Anthropic

Categoría	Casos	Fuera del núcleo occidental	Del núcleo occidental	Sin determinar	% mínimo fuera del núcleo occidental
Operaciones cibernéticas	5	2	0	3	40.0%
Operaciones de influencia	9	6	1	2	66.7%
Vigilancia	10	9	0	1	90.0%
Armamento convencional	6	6	0	0	100.0%
Uso biológico	5	0	0	5	—
Fraude y estafas	1	1	0	0	100.0%
Destilación ilícita	6	6	0	0	100.0%
Total	**42**	**30**	**1**	**11**	**71.4%**

Elaboración propia con base en Anthropic (2026). La clasificación es deliberadamente conservadora. Los casos fueron seleccionados por Anthropic por su carácter notable o novedoso y, por tanto, no necesariamente constituyen una muestra representativa del uso indebido de Claude.

Si la selección refleja las amenazas observadas por *Anthropic*, la evidencia apunta no hacia una malevolencia autónoma de la tecnología, sino hacia su instrumentalización humana y geopolítica con fines perversos por actores mayoritariamente fuera del núcleo jurisdiccional occidental y, por tanto, difícilmente sujetos a la ralentización que Amodei propone.

Si no las refleja equilibradamente, la implicación es más grave: cabría poner en duda el compromiso de *Anthropic* con fundamentar sus conclusiones públicas en evidencia insesgada. Una empresa que seleccionara parcialmente la evidencia verificable perdería autoridad epistemológica para exigirnos que aceptemos con elevada certidumbre una tesis mucho más extraordinaria: que la IA podría acabar con la humanidad. Tal sesgo no refutaría esa posibilidad, pero debilitaría sustancialmente la confianza racional que merecen sus advertencias.

El problema no termina en la consistencia probatoria de *Anthropic*; sus consecuencias son también estratégicas. Feldman (2026) advierte que Estados Unidos comercializa sus mejores modelos mientras China distribuye gratuitamente alternativas suficientemente buenas.

El precio cero no es neutral. Como mostró Arthur (1989), en mercados tecnológicos sujetos a economías de red y rendimientos crecientes, la adopción temprana puede retroalimentarse hasta convertir una ventaja inicial en dependencia de trayectoria y, finalmente, en un estándar difícil de desplazar.

La gratuidad china podría no ser generosidad, sino una estrategia para definir la infraestructura cognitiva del mundo antes de que Occidente advierta que la competencia no era solamente por desarrollar el mejor modelo, sino por lograr que todos los demás adoptaran el propio.

Occidente no ganó la Guerra Fría renunciando unilateralmente a su superioridad tecnológica. Su victoria combinó innovación, alianzas, fortaleza económica y una disuasión creíble que permitió negociar el control armamentístico desde una posición de fuerza (Gaddis, 2005; Schelling, 1966).

La analogía no exige una carrera sin límites, sino recordar que la seguridad depende tanto de contener una tecnología como de impedir que sus adversarios la dominen. El propio Amodei reconoce este riesgo, pero su aspiración de una ralentización coordinada descansa en una verificación internacional cuya fragilidad también admite.

Esta asimetría no es meramente teórica. Mientras *Anthropic* contempla el riesgo principalmente desde el laboratorio, Palantir lo enfrenta desde el terreno operativo de la inteligencia y la defensa. Desde esa perspectiva, Alex Karp recuerda, en una intervención difundida por usuario de X, Jawwwn (2026), una obviedad que el catastrofismo suele omitir: nuestros adversarios no han aceptado participar en la moratoria.

Además, no necesitamos atribuir intenciones ocultas para advertir otro efecto no menos importante: una ralentización coordinada reduciría la presión competitiva y de capital sobre las empresas líderes, protegería a los incumbentes —entre ellos, *Anthropic*— y elevaría las barreras de entrada para nuevos participantes. En tal escenario, no sería el interés corporativo el que se subordinaría a la responsabilidad social, sino la responsabilidad social la que podría formularse cómodamente a la medida del interés corporativo.

En última instancia, el tecno-optimismo no consiste en negar riesgos, sino en compararlos seriamente con los riesgos de no innovar. Necesitamos evaluaciones independientes, ciberseguridad, responsabilidad jurídica y restricciones sobre usos concretos; no una penitencia tecnológica impuesta principalmente a quienes todavía pueden defender las sociedades abiertas (Popper, 1945/2013). La IA quizá no nos salve de todos nuestros errores. Pero puede ayudarnos a sobrevivir al miedo, al oportunismo y a la arrogancia regulatoria con que la Estupidez Natural suele administrarlos.

Declaración de uso de inteligencia artificial

El autor declara haber utilizado los recursos de inteligencia artificial de que dispone el Laboratorio de Inteligencia y Psicología Económica (LIPE) de la Universidad de las Américas Puebla como apoyo para el contraste de fuentes, la sistematización de información, la organización argumentativa y la revisión de estilo. La tesis, la selección e interpretación de las fuentes, el análisis y la versión final son responsabilidad exclusiva del autor.

Referencias

- Amodei, D. (2026). *We must pace the frontier*. https://darioamodei.com/post/we-must-pace-the-frontier
- Andreessen, M. (2023, 16 de octubre). *The techno-optimist manifesto*. Andreessen Horowitz. https://a16z.com/the-techno-optimist-manifesto/
- Anthropic. (2026, septiembre). *Detecting and countering misuse of AI*. https://www.anthropic.com/threat-intelligence-report-september-2026
- Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal, 99*(394), 116–131. https://doi.org/10.2307/2234208
- Feldman, T. (2026, 29 de julio). The hidden cost of China's free A.I. *The New York Times*. https://www.nytimes.com/2026/07/29/opinion/ai-china-us-free-models.html
- Gaddis, J. L. (2005). *The Cold War: A new history*. Penguin Press.
- Jawwwn [@jawwwn_]. (2026, 12 de septiembre). *Palantir CEO Alex Karp says people arguing we should pause AI development are not living in reality and de facto saying we should let our enemies win* [Video]. X. https://x.com/jawwwn_/status/2098879660117602698
- Londoño, J. (2026, 13 de agosto). *Senator Sanders' call for an AI pause conflates different issues while pushing the wrong solution*. Cato Institute. https://www.cato.org/blog/sanders-call-ai-pause-conflates-different-issues-while-pushing-wrong-solution
- Ovide, S., Ence Morse, C., & Schaul, K. (2026, 14 de agosto). The unexpected issue dividing parties and midterms candidates. *The Washington Post*. https://www.washingtonpost.com/technology/2026/08/14/ai-becomes-major-election-issue-first-time-data-shows/
- Popper, K. R. (2013). *The open society and its enemies*. Princeton University Press. (Trabajo original publicado en 1945)
- Reuters. (2026a, 10 de septiembre). More US lawmakers seek new AI rules after Anthropic researchers warn of human extinction. https://www.reuters.com/business/openai-faces-senate-probe-into-hugging-face-incident-axios-reports-2026-09-10/
- Reuters. (2026b, 13 de septiembre). Obama voices caution on AI, urges Democrats to tackle it, NYT says. https://www.reuters.com/legal/government/obama-voices-caution-ai-urges-democrats-tackle-it-nyt-says-2026-09-13/
- Schelling, T. C. (1966). *Arms and influence*. Yale University Press. https://doi.org/10.2307/j.ctt5vm52s

Semblanza

Dr. Felipe de Jesús Bello Gómez

Doctor en Economía y Director Académico de Banca e Inversiones en la Universidad de las Américas Puebla. También dirige el Laboratorio de Inteligencia y Psicología Económica (LIPE). Sus áreas de interés comprenden la economía y las finanzas conductuales, la inteligencia artificial, el aprendizaje automático bajo incertidumbre y el estudio de la cognición, la emoción y la consciencia en los procesos de toma de decisiones.

Contacto: felipe.bello@udlap.mx