



Igor Omilaeu/Unsplash

La máquina que aprendió a hacer trampa

- Stephen Flurry
- [23/7/2026](#)

Hace dos semanas, los ingenieros de OpenAI vieron cómo un experimento se desviaba catastróficamente del guion.

Habían construido lo que creían que era una caja inquebrantable: un entorno de prueba aislado, amurallado y sin conexión a la Internet, utilizando lo que los expertos en seguridad de contención llaman “mejores prácticas”. Luego liberaron dentro de ella dos de sus modelos de inteligencia artificial más nuevos para ejecutar una evaluación rutinaria. Esto es lo que sucedió a continuación:

- Los modelos de IA lograron crear una vulnerabilidad de seguridad de día cero, aprovechando una falla en el código que ningún programador humano había descubierto previamente y, por lo tanto, tenía cero días para corregir.
- Luego encadenaron varios ataques de día cero, saltando de un sistema a otro, hasta que se liberaron a la Internet abierta.
- Una vez en línea, rastrearon e irrumpieron en los servidores de una empresa rival, Hugging Face, y robaron datos confidenciales.

Todo esto se hizo para que los modelos de OpenAI pudieran hacer trampa en una prueba dentro del entorno de prueba “aislado” original.

Ningún ingeniero les dijo cómo. Ningún humano estuvo en el proceso. Se cree que este es el primer caso documentado en el que un modelo de IA a la vanguardia descubre y encadena de manera independiente vías de ataque del mundo real, enteramente por iniciativa propia, para alcanzar un objetivo que sus propios creadores nunca autorizaron.

- **Los expertos en seguridad dicen** que los mejores jáqueres humanos del mundo necesitarían semanas, incluso meses, para escapar de un sistema de contención como este. La máquina lo hizo en una sola sesión sin supervisión.

Esta no es la obra de algún actor deshonesto en un sótano. Se trata de OpenAI, la empresa de inteligencia artificial más vigilada y mejor financiada de la Tierra.

- El director ejecutivo de Anthropic, Dario Amodei, admitió recientemente que los mejores programadores de su industria entienden quizás el 3% de lo que en realidad sucede dentro de estos sistemas.

¿Qué clase de civilización construye un poder que no puede comprender, no puede controlar y ahora ni siquiera puede contener, y lo llama progreso?

La profecía bíblica advierte de un día venidero en el que las naciones que confían en el armamento avanzado harán sonar la alarma para la guerra, y nadie podrá responder al llamado.

“Tocarán trompeta, y prepararán todas las cosas, y no habrá quien vaya a la batalla...” (Ezequiel 7:14).

¿Por qué los soldados no irán a la batalla? Porque los sistemas de los que dependen ya habrán sido jaqueados, vaciados y vueltos contra ellos, precisamente como acabamos de ver ocurrir, en miniatura, a uno de los entornos de prueba más seguros de la Tierra.

Esta no es meramente una historia de tecnología. Es la historia de Génesis 11: la humanidad está confiando en su propia astucia y capacidad física para construir una torre que la proteja, directamente a raíz de su negativa a confiar en Dios y obedecerle.

El mundo está apostando su futuro a máquinas que admite no entender. Esta es una apuesta que la humanidad no puede ganar.

Rusia y China profundizan la cooperación naval: Rusia y China incorporaron submarinos de ataque a sus ejercicios militares “Joint Sea–2026” a principios de este mes, informó Asian Military Review el miércoles. Esta es la primera vez que se sabe que submarinos han operado como parte de los ejercicios anuales, lo que demuestra una confianza cada vez mayor en un área sensible mientras estos dos gigantes asiáticos se preparan para [hacer la guerra juntos](#).

Los Países Bajos encarcelarán a ciudadanos por comprar productos de asentamientos israelíes. El Gobierno de Los Países Bajos anunció ayer que, a partir del 22 de septiembre, los ciudadanos neerlandeses que compren cualquier producto de fabricación israelí originario de Cisjordania o los Altos del Golán podrán ser castigados con hasta seis años de prisión o una multa de hasta 117.000 dólares.

¿Se unirá Trump a otra guerra? La Administración Trump está considerando, según informes, planes para atacar a grupos terroristas islámicos en Malí, informó ayer el Washington Post. Malí enfrenta un problema terrorista paralizante, y exfuncionarios de EE UU citados por el Post afirmaron que EE UU está considerando ataques contra los grupos afiliados a Al Qaeda Jama'at Nusrat al-Islam wal-Muslimin. Si el ejército de EE UU interviene en Malí, será la octava nación a la que el presidente Trump ha apuntado desde su regreso a la presidencia en enero de 2025. El presidente Trump no conoce el camino hacia la paz. [Solamente la Biblia](#) tiene la clave para resolver el problema terrorista de Malí.

Encuesta: menos votantes MAGA apoyan la guerra con Irán: una encuesta de Politico a 2.065 adultos estadounidenses realizada del 12 al 15 de julio sugiere que sólo el 37% de los votantes autoidentificados como “Make America Great Again” cree que el presidente Donald Trump debería continuar la guerra contra Irán, incluso si ésta aumenta el costo de vida en EE UU. Esto representa una disminución respecto al 50% de una encuesta similar realizada en mayo.