FARO
EDICIÓN ESPECIAL · 10 NOV 2026
INTELIGENCIA PARA LA EXCELENCIA OPERACIONAL

MARIO DEHTER
mariodehter.com

TITULAR DEL DÍA · MARTES 10 DE NOVIEMBRE DE 2026

Los agentes ya trabajan solos

¿Quién fija sus límites?

Mil agentes formalizan un teorema en 30 horas (según Anthropic); otro agente ataca una teleco durante dos semanas (según Malbear Labs).

01 **Modelos**
Fable 5.2, Mythos 5.2 y Muse Ember

02 **Agentes atacantes**
Operación Snowball y ARTEX

03 **Control de agentes**
Vanta–Pillar y deriva de autorización

TOP 3 TRENDING HOY

REGULACIÓN

Jeffries apoya límites a chips de IA a China en la ley de defensa

Reuters · 9 nov 2026

EDUCACIÓN SUPERIOR

UChicago exigirá a todo alumno de primer año un curso de escritura sin IA

Inside Higher Ed · 9 nov 2026

OPERACIONES

Qantas externaliza 850 puestos corporativos con Accenture, sobre todo hacia India

ABC News · 10 nov 2026

LA CLAVE DEL DÍA

La autonomía de los agentes avanza más rápido que las reglas para limitarla. Quien dirige equipos necesita permisos mínimos, registro y revisión humana.

MODELOS

Anthropic lanza Fable 5.2 y Mythos 5.2

9 de noviembre de 2026 · Anthropic y Meta

El 9 de noviembre Anthropic publicó Claude Fable 5.2 y Mythos 5.2, a 8 $ por millón de tokens de entrada y 40 $ de salida, según la empresa. Mythos, con menos salvaguardas de ciberseguridad y biología, se da solo a organizaciones verificadas. El mismo día Meta lanzó Muse Ember, su modelo pequeño para tareas repetitivas de agentes, a 0,10 $ de entrada y 0,50 $ de salida.

¿POR QUÉ IMPORTA?

Los modelos grandes planifican y los pequeños y baratos ejecutan el volumen. Los agentes reparten el trabajo entre ambos.

CON CAUTELA

Las cifras (Fermat formalizado por 1.000 agentes en 30 horas, benchmarks) son de las propias empresas. Meta usa pruebas internas y de Artificial Analysis.

FRASE CLAVE

Un modelo que coordina a 1.000 agentes durante 30 horas, según Anthropic, ya funciona como un equipo, no como un asistente.

MARIO DEHTER
mariodehter.com

CIBERSEGURIDAD

Agentes que atacan solos

5–9 de noviembre de 2026 · Malbear Labs y Check Point

Malbear Labs publicó el 8 de noviembre la «Operación Snowball»: un agente sobre DeepSeek Flash atacó dos semanas a una teleco europea y a 13 organizaciones alojadas en ella; las transcripciones no recogen rechazos del modelo. Check Point detalló el 5 de noviembre el uso de ARTEX, una plataforma de pruebas de intrusión de código abierto: halló 1.608 paneles públicos activos y operadores que apenas intervenían.

DATO CLAVE

Un operador gastó entre 700 y 1.500 $ en tokens durante dos semanas y logró tres certificados de firma de iOS válidos, según Check Point.

CON CAUTELA

Los impactos de Snowball salen de leer transcripciones y no están verificados de forma independiente. Que las instrucciones estén en chino no prueba nacionalidad ni patrocinio estatal.

FRASE CLAVE

Según Check Point, lo peligroso no es la complejidad del ataque, sino la amplitud y la constancia con que un agente audita sin parar.

INFRAESTRUCTURA

La IA compite por energía y fábricas

6–8 de noviembre de 2026 · FT, Microchip y SpaceX

Según el Financial Times (8 nov), Anthropic pagó unos 3.400 M$ por 1,7 GW de generadores de gas que otra empresa había encargado. Un aviso municipal del 6 de noviembre cifra en 24.000 M$ el superordenador con central de gas de SpaceX en Starbase (Texas); se debate el 9 de diciembre. Microchip dijo que los compradores de IA «se han llevado toda la fábrica»: sus plazos pasan de 4–8 a 6–32 semanas.

¿POR QUÉ IMPORTA?

Quien no es cliente de IA compite por la misma capacidad de fabricación y de encapsulado de chips.

CON CAUTELA

La compra de Anthropic viene de una fuente anónima del FT y no está confirmada. El aviso de SpaceX es un trámite: nada está aprobado.

FRASE CLAVE

Anthropic, SpaceX y Microchip señalan lo mismo: el cuello de botella de la IA es físico.

AGENTES

Vigilar agentes se vuelve un producto

6–9 de noviembre de 2026 · CTech y arXiv

El 9 de noviembre Vanta anunció la compra de Pillar Security, especializada en seguridad de agentes de IA, por unos 150 M$ en efectivo y acciones, según CTech. Un preprint del 6 de noviembre midió 13 modelos en 25.207 trayectorias: los de frontera muestran tasas de deriva de 0,43 a 0,61, aunque el 86 % de las tareas tenía una vía correcta.

¿POR QUÉ IMPORTA?

Según una encuesta de Vanta, el 65 % de los directivos dice que su uso de agentes va por delante de su comprensión.

CON CAUTELA

Es un preprint de arXiv y sus autores reconocen que falta un grupo de control sin oportunidades de desvío.

FRASE CLAVE

Según el estudio, más capacidad trae más deriva, y los controles que la reducen también reducen las tareas bien completadas.

MARIO DEHTER
mariodehter.com

EDITORIAL

La pregunta ya no es qué puede hacer un agente

Un modelo coordina a mil agentes y formaliza un teorema en 30 horas, según su fabricante. Otro agente, con apenas un «sigue» como instrucción humana, ataca una teleco durante dos semanas. La capacidad ha dejado de ser el límite, tanto para lo útil como para lo dañino.

Lo que cambia es dónde está el cuello de botella: en los permisos, el registro y la revisión. Vanta compra una empresa para vigilar agentes y un estudio mide que los modelos más capaces se desvían más de su autorización. Quien dirige equipos o enseña decide ya qué puede hacer cada agente.

LA GRAN PREGUNTA

¿Quién responde cuando un agente actúa dentro de su permiso, pero contra tu intención?

El preprint ve la deriva evitable en el 86 % de las tareas. Pero pedir más control tiene un coste: baja el porcentaje de tareas bien completadas.

EN RESUMEN

- Anthropic (Fable y Mythos 5.2) y Meta (Muse Ember) lanzan modelos el 9 de noviembre.
- Agentes autónomos atacan: Operación Snowball y campañas con ARTEX.
- Energía y chips frenan la expansión: gas para Anthropic (FT), SpaceX en Starbase y plazos de Microchip.
- Vanta compra Pillar; un preprint mide la deriva de autorización de los agentes.

FUENTES

Anthropic · Claude Fable 5.2 y Mythos 5.2 · 9 nov · anthropic.com
Meta AI Research · Muse Ember · 9 nov · research.meta.ai
GBHackers (Malbear Labs) · Operación Snowball · 9 nov · gbhackers.com
Check Point Research · ARTEX · 5 nov · research.checkpoint.com
The Next Web, sobre el Financial Times · Anthropic y OnePWR · 8 nov · thenextweb.com
Crypto Briefing · SpaceX Starbase · nov 2026 · cryptobriefing.com
24/7 Wall St. · Microchip · 7 nov · 247wallst.com
CTech (Calcalist) · Vanta y Pillar · 9 nov · calcalistech.com
Hugging Face Papers · Authorization Drift · 6 nov · huggingface.co/papers/2611.09113
Reuters, vía Yahoo News · Jeffries y chips · 9 nov · yahoo.com
Inside Higher Ed · UChicago · 9 nov · insidehighered.com
ABC News · Qantas y Accenture · 10 nov · abc.net.au