EL PARADIGMA QUE CAMBIÓ

Mythos: la nueva era de las amenazas IA-agénticas.

Anthropic anunció **Mythos**, un modelo frontera con capacidades excepcionales para **descubrir y explotar vulnerabilidades de manera autónoma**, a velocidad de máquina. La industria reaccionó con titulares apocalípticos. Pero la evidencia técnica cuenta una historia más matizada —y más optimista para quien sepa leerla.

Lo que ya pasó. Lo que viene.

El modelo —que aún no es de acceso general— fue probado bajo **Project Glasswing**, una alianza de Anthropic con AWS, Apple, Cisco, CrowdStrike, Google, JPMorgan, Microsoft, NVIDIA y Palo Alto Networks. Los primeros resultados son contundentes:

271
zero-days encontrados en una sola versión de Firefox 150

+4.000
vulnerabilidades proyectadas para próximas auditorías de Windows

Horas
la nueva ventana entre publicación de parche y exploit funcional

La ventana entre que un parche se publica y un exploit funcional aparece **se redujo de semanas a horas**. La habilidad necesaria para ejecutar un ataque serio sigue cayendo. Para muchas organizaciones, eso significa repensar cómo operan sus equipos de seguridad —no incrementar el pánico.

El doomerism vende titulares. El contexto gana batallas.

La narrativa del pánico asume que **capacidad del modelo = ventaja del atacante**. La evidencia muestra otra cosa. Aikido Security corrió **1.000 pentests con IA** en condiciones controladas. El resultado:

Cuando la IA tiene contexto (acceso al código fuente, runtime, dependencias), encuentra 7× más vulnerabilidades críticas y opera al doble de eficiencia.

WHITEBOX VS. GREYBOX · AIKIDO AI PENTESTING STUDY · ABRIL 2026

El factor decisivo no es la capacidad del modelo: es el contexto. Y el contexto vive con el defensor. Tú tienes el código, el runtime, el árbol de dependencias y la lógica de negocio. El atacante opera a ciegas desde fuera. Mythos no inclinó la balanza hacia los atacantes —cambió el ritmo, sí, pero la ventaja estructural sigue del lado del defensor que sabe usarla.

Globaltek

CÓMO PREPARARTE

Postura Mythos-Ready

5 PILARES ACCIONABLES

Construye una postura Mythos-Ready.

Estos cinco frentes no son aspiracionales: son las preguntas concretas que tu organización debería poder responder con un "sí" antes de que la próxima oleada de modelos frontera salga al público.

01 · INVENTARIO

Conoce lo que corres

Inventario continuo de tu superficie de ataque. Incluye **supply chain agéntica** (MCP servers, plugins, agentes) y aplicaciones generadas con vibe coding (Lovable, Bolt, Replit).

02 · BLAST RADIUS

Reduce lo que puede fallar

MFA con llave hardware para accesos privilegiados. Segmentación de ambientes (dev/QA/prod) en cuentas cloud separadas. Scoped tokens para agentes y CI/CD.

03 · SUPPLY CHAIN

Controla tu cadena de software

Usa **lockfiles** contra slopsquatting. Escanea dependencias en busca de malware. Despliega EDR en las máquinas de tus desarrolladores —son objetivo de alto valor.

04 · ENCUENTRA PRIMERO

Atácate antes que el atacante

Pentest con IA sobre tu propio código. AI SAST como práctica continua. Gate de revisión obligatorio para código generado por IA antes de llegar a producción.

05 · PARCHA RÁPIDO Y OBSERVA

Opera a velocidad de máquina

Logging **tamper-resistant off-host**. Monitoreo de tráfico saliente y comportamiento de cuentas de servicio. SLA objetivo: menos de **24 horas** entre parche crítico disponible y código corriendo en producción.

Globaltek opera a la velocidad de la amenaza.

CsRA
Ciberseguridad Resiliente Automatizada. Detección, contención y respuesta a velocidad de máquina.

Pentest IA
Validamos tu postura con la misma tecnología que usan los atacantes —pero con tu contexto.

Observabilidad
Logging tamper-resistant, monitoreo de salida y detección de comportamiento anómalo.

comercial@globalteksecurity.com
www.globaltek.co · ISO 9001 · ISO 27001

AGENDA TU DIAGNÓSTICO

FUENTES

aikido

Investigación basada en publicaciones de **Aikido Security** (Sooraj Shah · Dania Durnas · Mike Wilkes, abril-mayo 2026), **Anthropic Mythos Preview System Card**, **Project Glasswing** y el reporte de Firefox 150 (Ars Technica · Mozilla).