Mythos: la Amenaza Fantasma

April 19, 2026

Hace algunos días fui invitado a Radio Pauta de Chile para analizar un tema que, hasta hace muy poco, parecía propio de la ciencia ficción: la aparición de una nueva red neuronal avanzada desarrollada por la empresa Anthropic, conocida como Claude Mythos. Lejos de ser un modelo más dentro del ecosistema de inteligencia artificial (que ya es monstruoso), todo indica que estamos frente a un **punto de inflexión** que podría redefinir no solo la ciberseguridad, **sino la estabilidad de sistemas críticos a nivel global**.

Para llegar a esa conversación con fundamentos sólidos, dediqué varios días a investigar en profundidad, revisando informes técnicos, análisis académicos y cobertura especializada en medios internacionales. A medida que avanzaba, el patrón se volvía cada vez más claro y, al mismo tiempo, más inquietante. No se trataba únicamente de un avance en capacidad computacional o eficiencia, sino de una nueva forma de inteligencia capaz de entender sistemas complejos, encontrar debilidades ocultas y conectar puntos que durante años han pasado desapercibidos incluso para expertos humanos. Lo que realmente llama la atención no es solo lo que Claude Mythos puede hacer, sino el impacto que podría tener en sectores altamente sensibles. Entre ellos, **el sistema financiero mundial aparece como uno de los más expuestos.**

Bancos, infraestructuras de pago, sistemas de crédito y redes de datos concentran exactamente las condiciones que este tipo de inteligencia puede analizar, mapear y eventualmente vulnerar. Las advertencias ya no vienen únicamente de laboratorios tecnológicos, sino también de expertos en ciberseguridad (ver "Gummo" en YouTube) y actores del propio sistema financiero, que comienzan a dimensionar el alcance del riesgo.

Este artículo recoge los principales hallazgos de esa investigación. Más que una descripción técnica, es una mirada estratégica sobre lo que está en juego: una nueva generación de inteligencia artificial que no solo responde preguntas, sino que puede descubrir cómo funcionan, y cómo fallan, los sistemas más críticos del mundo.

La aparición de “Claude Mythos” cambia por completo la forma en que se descubren y explotan las vulnerabilidades en los sistemas digitales. Antes, la inteligencia artificial ayudaba a los expertos humanos, pero ahora estamos frente a algo distinto: **un sistema capaz de analizar por sí solo, entender cómo funcionan los sistemas y encontrar fallas de manera profunda**.
No se trata solo de hacerlo más rápido, sino de hacerlo mejor, entendiendo el sistema completo. Esto convierte la búsqueda de vulnerabilidades en un proceso constante y mucho más potente que antes. Algo clave es que Mythos no fue creado solo para ciberseguridad. Es una inteligencia general que, gracias a su capacidad para entender código y sistemas complejos, termina siendo muy buena encontrando fallas. Esto es especialmente importante en los bancos, donde los sistemas son muy complejos, mezclan tecnologías nuevas con antiguas y dependen de muchas conexiones con otros sistemas. Una IA como Mythos puede ver todo eso junto, no por partes, y detectar problemas que antes pasaban desapercibidos.

Aquí aparece el concepto más importante: “mythos”. No significa solo encontrar errores, sino entender la historia completa de cómo un sistema puede fallar. Es decir, no ve una falla aislada, sino cómo una pequeña debilidad puede convertirse en un problema grande al combinarse con otras. Por ejemplo, puede detectar cómo un error en una página web termina permitiendo el acceso al sistema operativo. Esa capacidad de conectar puntos es lo que permite encontrar vulnerabilidades que han estado ocultas durante años. Los expertos en seguridad coinciden en que esto cambia totalmente el ritmo del juego. Antes, una vulnerabilidad podía tardar semanas o meses en ser explotada. Ahora, ese tiempo puede reducirse a minutos. Y

esto no es algo del futuro, ya está pasando con herramientas actuales basadas en inteligencia artificial.

En la práctica, significa que las organizaciones ya no compiten solo contra personas, sino contra sistemas automatizados que prueban miles de formas de ataque en muy poco tiempo.

En el caso de los bancos, el problema es aún más serio. Las instituciones financieras tienen justo lo que estos sistemas buscan: mucha información sensible, sistemas antiguos difíciles de mantener, estructuras complejas de acceso y cada vez más uso de inteligencia artificial en procesos importantes como fraudes, créditos o atención al cliente. Todo eso crea un entorno donde hay muchas posibles fallas, y ahora esas fallas pueden ser encontradas mucho más fácilmente. Además, Mythos no solo encuentra errores en el código. También detecta problemas en cómo están diseñados los sistemas. Por ejemplo, puede identificar fallas en cómo se gestionan los accesos, cómo se comunican las máquinas entre sí o cómo se organizan las autorizaciones. En un banco, esto es crítico, porque si alguien logra entrar por una de esas fallas, puede moverse dentro del sistema y acceder a información muy sensible sin ser detectado.

Otro punto importante es que esta tecnología tiene dos caras. Sirve para defender, pero también para atacar. Lo que ayuda a encontrar vulnerabilidades para corregirlas, también puede ser usado por atacantes para explotarlas. Y según los expertos, esto no es una posibilidad futura, ya está ocurriendo. La gran preocupación es qué pasará cuando estas capacidades estén disponibles de forma más amplia y sean usadas por más actores. A esto se suma un riesgo adicional: los propios bancos están usando inteligencia artificial en sus procesos.

Eso significa que no solo sus sistemas tradicionales pueden ser atacados, sino también sus sistemas de IA. Estos pueden ser manipulados para filtrar información, tomar malas decisiones o actuar fuera de control si no están bien protegidos. En otras palabras, la IA no solo ayuda, también puede convertirse en una nueva puerta de entrada para ataques.

En el fondo, lo que cambia con Mythos no es solo la tecnología, sino la forma de entender la seguridad. Antes se trabajaba con amenazas conocidas y controles definidos. Ahora estamos entrando en un escenario donde las vulnerabilidades se descubren constantemente, gracias a sistemas que "razonan" de manera automática.

Esto obliga a cambiar la forma de proteger los sistemas, especialmente en sectores como el bancario donde la confianza es fundamental.

La conclusión es clara: el mayor impacto de Mythos no es una vulnerabilidad específica, sino su capacidad de encontrar muchas y de entender cómo se conectan entre sí. En un mundo donde los sistemas son cada vez más complejos, lo importante ya no es solo proteger cada parte, sino entender cómo todo puede fallar en conjunto. Para los bancos, esto significa que la seguridad ya no puede ser reactiva. Tiene que ser continua, adaptable y al mismo nivel de velocidad e inteligencia que las nuevas amenazas impulsadas por la inteligencia artificial.