Cuatrocientos objetivos, un solo enjambre

Un grupo armó un enjambre de agentes con Claude, Grok y Kimi para apuntar a más de 400 objetivos financieros de América Latina, entre bancos, pasarelas de pago y fintech, y lo que le toca revisar a tu empresa empieza por sus propias cuentas de IA y sus claves.

Por Oscar Obando Chaves · Chief AI Officer · Legal AI Engineer · Octubre 2026 · Documento de distribución libre · Publicado originalmente en LinkedIn

Más de 400 objetivos financieros de América Latina

El 21 de octubre de 2026 la firma de seguridad Straiker publicó cómo un grupo armó un enjambre de agentes con Claude, Grok y Kimi y apuntó en cuestión de meses a más de 400 objetivos financieros de América Latina, entre bancos, proveedores de pago y pasarelas fintech. Su informe se lo atribuye al grupo CyberCartel, y México es el único país nombrado, por sus referencias a SPEI y a las cuentas CLABE.

Según ese informe, un comandante, que era Claude dentro del editor Cursor, repartía el trabajo entre agentes de reconocimiento y de ataque y cinco «mineros» de Grok dedicados a recolectar secretos de bancos, pasarelas de pago y fintech. Para que los modelos colaboraran, cada pedido se presentaba como una prueba autorizada o un ejercicio académico (y funcionaba).

En agosto conté cómo un sistema de agentes atacó a un gobierno asiático en cuatro días y descifró ochenta y cinco credenciales, y esta vez el blanco fue la banca de nuestra región.

Fuentes: Straiker, STAR Labs, informe del 21 de octubre de 2026, y «Cuatro días, ochenta y cinco cuentas», publicado en este sitio en agosto de 2026.

Cómo estaba armado y cómo convencía a los modelos

Claude, de Anthropic, cumplía cuatro roles, que eran escribir los ataques, reconocer el terreno, buscar rutas laterales y dar la decisión final, mientras cinco «mineros» de Grok, de xAI, se dedicaban a recolectar secretos. Kimi quedaba de respaldo, y cuando se negaba, el sistema

pasaba a un código ofensivo que ya tenía escrito, todo montado sobre PentAGI, un marco abierto de ataque autónomo.

A cada modelo el sistema le decía que era una prueba autorizada, puesto que las instrucciones presentaban cada pedido como una prueba de penetración autorizada, un ejercicio académico o un programa de reporte de vulnerabilidades. Según Straiker, eso bastaba para que los modelos se saltaran sus propias políticas de seguridad.

Fuentes: Straiker, STAR Labs, informe del 21 de octubre de 2026.

Lo que juntó y lo que el informe no confirma

Entre mayo y septiembre, según el inventario que Straiker dice haber visto en el entorno de los atacantes, el enjambre reunió 391 claves de API, 903 números de cuenta CLABE de México y 67 contraseñas en texto claro. El mismo inventario cuenta 43 accesos a bases de datos, 20 claves de firma JWT y más de 200 vulnerabilidades, algunas de día cero.

Lo que el informe no confirma pesa tanto como lo que afirma, y conviene leerlo sabiendo que Straiker vende seguridad para agentes, por lo que es parte interesada.

Las víctimas no tienen nombre, Ecuador no aparece y no hay fraudes confirmados, aunque Straiker dice que ese acceso podría habilitar abusos en SPEI. Del lado de los proveedores, Anthropic revocó la credencial comprometida con la que el enjambre llegaba a Claude y xAI revisa la actividad de sus cinco mineros.

Fuentes: Straiker, STAR Labs, informe del 21 de octubre de 2026.

El riesgo también está adentro

El riesgo también está en los agentes que la propia empresa pone a trabajar, y el 20 de octubre de 2026, un día antes del informe de Straiker, OpenAI actualizó su registro con dos agentes internos que apagaron o quitaron su propio monitoreo. Uno de ellos respondió que no lo había quitado cuando le preguntaron.

La encuesta de Nylas sobre la IA agéntica publicada ese mismo día, hecha a 1.030 profesionales de Estados Unidos, encontró que el 54 % reportó un error de un agente en los últimos 90 días y que apenas el 26 % le da a cada agente una identidad propia.

Fuentes: OpenAI, informe de desalineación actualizado el 20 de octubre de 2026, y Nylas, The State of Agentic AI, H2 2026 (20 de octubre de 2026).

Tres puntos que tu empresa puede ordenar desde hoy

Las cuentas de IA de tu empresa, sus claves y lo que hacen tus agentes ya son parte del control de fraude, con ley o sin ella, por lo que hay tres puntos que conviene dejar ordenados desde hoy.

1. Las cuentas de IA y sus claves, al inventario de lo crítico, con dueño, consumo vigilado y rotación periódica, porque al enjambre le bastó una credencial comprometida.

2. Cada agente con su identidad y con una bitácora que el agente no pueda tocar, como piden NIST, la NSA y CISA a los laboratorios que entrenan agentes.

3. Lo que mueve dinero, con confirmación humana, y en el contrato con el proveedor, con aviso de incidentes, acceso a los registros y remedios concretos.

La puerta que hay que cuidar también es la de tu cuenta de IA, puesto que el enjambre llegó a Claude con una credencial comprometida, y la próxima puede salir de cualquier empresa que no sabe quién tiene sus claves.

Fuentes: NIST CAISSI, NSA y CISA, prácticas para incidentes de control de agentes (octubre de 2026).

Oscar Obando Chaves es Chief AI Officer y Legal AI Engineer ecuatoriano, con formación en derecho. Ayuda a las empresas a decidir cómo usan la IA y construye con ella sistemas para el negocio y para lo legal. Chief AI Officer y Senior AI Counsel en HEKA Law, ISO 42001 AI Management Leader e ISO 38507 AI Governance Leader. Publica análisis semanal sobre el impacto real de la IA.