Google reveló que hackers vinculados a China, Rusia y Corea del Norte usan Gemini para acelerar ciberataques

15/02/2026

Un informe del equipo de inteligencia de amenazas de Google reveló que **hackers vinculados a gobiernos** utilizaron **Gemini** para **agilizar** distintas fases de **ciberataque**.

Elaborada por el **Google Threat Intelligence Group**, la investigación afirmó que el **uso de la herramienta de IA**no se limitó a campañas básicas de phishing, sino que incluyó **tareas de reconocimiento**, redacción de **ingeniería social**, asistencia en programación y actividades posteriores a una intrusión.

Según el reporte, la actividad observada involucra **células asociadas a China, Irán, Corea del Norte y Rusia.** Los prompts y respuestas analizados incluyeron perfilado de objetivos,

generación de textos persuasivos para engaños, **traducciones, ayuda con código**, pruebas de vulnerabilidades y depuración de herramientas cuando fallaban durante una intrusión.

Google reveló que hackers vinculados a China, Rusia y Corea del Norte usan Gemini para acelerar ciberataques. (Imagen: GeminiAI)

Los investigadores de Google plantean una idea central: **la IA no introduce tácticas radicalmente nuevas, sino que acelera procesos** que los atacantes ya realizaban. El reconocimiento previo a un ataque, la creación de señuelos creíbles, la modificación de herramientas y la corrección de errores técnicos forman parte del manual habitual de operaciones ofensivas.

La diferencia está en el ritmo. Con asistencia de modelos como Gemini, los **hackers** pueden **obtener reescrituras rápidas, soporte multilingüe o correcciones de código en cuestión de segundos**. Eso reduce fricciones y acorta los ciclos entre prueba y error.

En uno de los casos descritos, vinculado a actores asociados a China, un ciberdelincuente adoptó la identidad de un experto en seguridad informática dentro de un escenario ficticio para

solicitar la **automatización del análisis de vulnerabilidades** y la generación de planes de prueba dirigidos. En otro ejemplo, un hacker, también basado en el país asiático, recurrió varias veces al modelo para **tareas de depuración**, investigación técnica y orientación relacionada con intrusiones.

El riesgo está en la velocidad de la IA

Uno de los puntos que subraya Google es el **impacto en los tiempos**. Si los grupos pueden interactuar **más rápido** sobre objetivos y herramientas, los equipos de defensa disponen de menos margen entre las primeras señales y el daño efectivo.

Menos tiempo también implica **menos pausas** visibles en los registros, menos errores manuales y menos repeticiones que podrían delatar la actividad. La **automatización** parcial, incluso en tareas rutinarias, puede **alterar la dinámica entre atacante y defensor.**

El informe también advierte sobre otra práctica distinta al uso operativo directo: la extracción de modelos y la destilación de conocimiento. En estos escenarios, actores con acceso autorizado a APIs realizan grandes volúmenes de consultas para replicar el comportamiento y la lógica del sistema, con el objetivo de entrenar otro modelo.

Google menciona un caso que involucró alrededor de **100.000 prompts** orientados a **reproducir comportamientos** en tareas en idiomas distintos del inglés. Según el informe, este tipo de actividad puede generar daños comerciales y de propiedad intelectual si se escala.

Qué deberían vigilar los equipos de

seguridad

Google informó que deshabilitó cuentas e infraestructura vinculadas al abuso documentado y que incorporó defensas específicas en Gemini. También indicó que continúa probando y reforzando sus mecanismos de seguridad.

Para los equipos de **ciberseguridad**, **la conclusión práctica es asumir que los ataques asistidos por IA avanzarán más rápido**, aunque no necesariamente sean más sofisticados. Conviene prestar atención a mejoras repentinas en la calidad de los señuelos, ciclos más veloces en el desarrollo de herramientas y patrones inusuales de uso de API.

Fuente: TN