LITORALPRESS

Fecha: 15-01-2026
Medio: El Mercurio
Supl. : El Mercurio - Cuerpo B
Tipo: Noticia general
Título: Cómo la IA está facilitando la vida a los ciberdelincuentes

Pág. : 6
Cm2: 811,7

Tiraje: 126.654
Lectoría: 320.543
Favorabilidad: No Definida

Estafas digitales más eficientes:

Cómo la IA está facilitando la vida a los ciberdelincuentes

El *phishing* y otras ciberestafas son cada vez de mayor envergadura, más específicas y más convincentes.

WILLIAM BOSTON
THE WALL STREET JOURNAL

Gracias a los rápidos avances en la inteligencia artificial (IA), los ciberdelincuentes que buscan engañarlo para que les entregue sus fondos de jubilación o les revele secretos de la empresa se están volviendo más audaces y más fuertes.

Del mismo modo en que la IA puede personalizar los avisos que ve en línea, los actores malvados la están utilizando para obtener información personal que les permite crear estafas personalizadas en forma rápida y a gran escala. Las compañías de IA, como Anthropic, OpenAI y Google, afirman que los ciberdelincuentes están utilizando su tecnología para realizar detallados esquemas de *phishing* (estafa digital), crear *malware* (programas o códigos maliciosos) y llevar a cabo otros ciberataques. Expertos aseguran que herramientas de IA similares se están utilizando para crear *deepfake* (ultrafalsificación) de audios y videos de ejecutivos corporativos para tratar de arrancar información a los empleados en forma inadvertida.

Las corporaciones y las entidades de gobierno pronto podrían enfrentarse a ejércitos de agentes de IA que están aprendiendo cómo identificar vulnerabilidades en redes computacionales, luego planear y ejecutar un ataque casi sin ninguna intervención humana en absoluto.

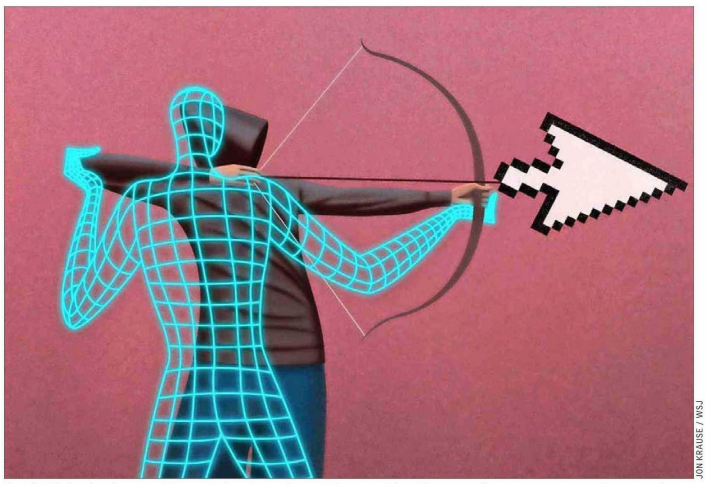
¿Cómo están utilizando la IA los delincuentes? ¿Qué tan avanzados y autónomos han llegado a ser sus ciberbots?

A continuación, las respuestas a estas y otras preguntas.

¿Cómo la IA está cambiando las capacidades de los ciberdelincuentes?

La IA está haciendo que los ciberdelincuentes sean más eficientes, permitiéndoles ampliar las operaciones. Anthropic, la compañía que está detrás del agente de IA, Claude, indica que esta tecnología puede amplificar la velocidad, el alcance y la automatización de los ataques. "El cambio real está en el alcance y la escala", señala Alice Marwick, directora de investigación en Data & Society, un instituto de estudios tecnológicos independiente sin fines de lucro. "Las estafas son más grandes, más específicas, más convincentes".

Entre la mitad y las tres cuartas partes del *spam* y el *phishing* en el mundo ahora se generan con IA, afirma Brian Singer, candidato a un doctorado en la Universidad Carnegie Mellon quien investiga el uso de grandes modelos de lenguaje para ciberataques y defensas.



JON KRAUSE / WSJ

Con la IA, los hackers podrían examinar las redes sociales para identificar aquellas personas que están pasando por grandes cambios en su vida y que por lo mismo serían vulnerables a caer en una estafa virtual.

¿Cuáles son algunos ejemplos?

Los ataques de *phishing* se están volviendo más creíbles. Por ejemplo, la IA capacitada en comunicaciones de una compañía ahora puede redactar miles de mensajes fluidos y sobre la marca que imitan el tono de un ejecutivo o mencionan eventos actuales sacados de datos públicos.

La IA también se puede utilizar para mejorar la gramática y el lenguaje, lo que ayuda a los estafadores extranjeros a superar las barreras lingüísticas que podrían haber hecho que sus intentos de *phishing* parecieran menos creíbles en el pasado. Igualmente la IA puede ayudar a los delincuentes a personificar a otros a través de *deepfakes* y clonación de voz, e incluso utilizar el mismo personaje falso para atacar a múltiples personas.

El cambio más grande es "la credibilidad a escala", afirma John Hultquist, analista jefe de Google Threat Intelligence Group.

Los delincuentes también están mejorando en cuanto a encontrar objetivos vulnerables. Por ejemplo, pueden desplegar la IA para examinar las redes sociales con el fin de identificar personas que están pasando por grandes cambios en su vida —un divorcio, una muerte en la familia, la pérdida del empleo— que podrían dejarlas más vulnerables a caer en una estafa sentimental, de inversión o laboral.

¿La IA ha facilitado que las personas cometan ciberdelitos?

Totalmente. Ahora hay mercados en la *dark web* en los que las personas con menos conocimientos tecnológicos pueden arrendar o comprar herramientas de IA para operaciones criminales por apenas US$ 90 al mes. "Los creadores venden suscripciones para atacar plataformas con precios escalonados y soporte al cliente", señala Nicolas Christin, jefe del departamento de software y sistemas sociales de Carnegie Mellon. Con nombres como WormGPT, FraudGPT y DarkGPT, estas herramientas se pueden utilizar para crear *malware* y campañas de *phishing*, e incluso ofrecen tutoriales de hackeo.

"No necesita saber cómo codificar, sino solo dónde encontrar la herramienta", indica Margaret Cunningham, vicepresidenta de seguridad y estrategia de IA en Darktrace, una firma de ciberseguridad.

Una tendencia más nueva, llamada *vibe-coding* o *vibe-hacking*, podría permitir que los ciberdelincuentes aficionados utilicen la IA para desarrollar sus propios programas maliciosos en lugar de comprarlos en la *dark web*. Anthropic reveló a principios de este año que había frustrado varias instancias en que "delincuentes con pocas habilidades técnicas" habían utilizado su IA, Claude, para crear *ransomware* (secuestro de datos).

¿Cómo la IA está cambiando las organizaciones criminales?

Incluso antes del surgimiento de la IA, la ciberdelincuencia operaba como un mercado, afirman expertos. Un ataque característico de *ransomware* involucraba actores separados: agentes de acceso quienes se introducían en redes corporativas y vendían puntos de ingreso, equipos de intrusión que se movían a través de sistemas y robaban datos, y operadores de *ransomware* como un servicio que desplegaban el *malware*, manejaban negociaciones y dividían las ganancias.

Lo que ha cambiado con la IA es la velocidad, la escala y la accesibilidad de ese ecosistema. Las tareas que antes hacían humanos y que requerían habilidades técnicas ahora se pueden automatizar, permitiendo que estas organizaciones disminuyan su tamaño, minimicen el riesgo y maximicen las ganancias. "Piense en esto como el siguiente nivel de la industrialización. La IA aumenta el rendimiento sin que se requiera mano de obra más especializada", explica Christin, el profesor de Carnegie Mellon.

¿Puede la IA lanzar un ciberataque en forma autónoma?

La respuesta corta: no todavía. Los expertos lo comparan con la carrera por construir automóviles totalmente autónomos. El primer 95% se ha logrado, pero el último tramo, que permitiría que el vehículo se condujera por sí solo en forma confiable en cualquier parte y en cualquier momento, sigue siendo escurridizo. Los investigadores están probando las capacidades de hackeo de la IA en laboratorio, y un equipo de Carnegie Mellon, con el respaldo de Anthropic, a principios de este año duplicó la malvada violación de datos de Equifax mediante IA. "Es un gran salto", asegura Singer, el candidato a un doctorado quien encabezó el proyecto del Instituto de Seguridad y Privacidad CyLab de Carnegie Mellon.

EFECTO IA
En el mundo, entre la mitad y las tres cuartas partes del *spam* y el *phishing* se generan con IA.

Aunque los investigadores como Singer han demostrado que la IA es capaz de planificar y llevar a cabo un ataque por su cuenta en un laboratorio, una mayoría de expertos no cree que la tecnología haya avanzado hasta ese punto en el mundo real. Sin embargo, Anthropic hace poco reveló que su IA, Claude, se utilizó para realizar un ataque casi por su cuenta, lo que sugiere que la IA se está acercando a la autonomía.

"Dentro de dos o tres años, la ciberseguridad será algo como IA versus IA, porque los humanos no podrán ir al mismo paso", observa Singer.

¿Necesitamos cambiar la forma en que nos defendemos?

Los *hackers* pueden estar aprovechando la IA para cometer delitos, pero las compañías de IA afirman que la misma tecnología también puede ser utilizada por las organizaciones para reforzar sus ciberdefensas. Llámela la nueva carrera armamentista de la IA. Anthropic y OpenAI, por ejemplo, están desarrollando modelos de IA que pueden inspeccionar en forma autónoma y constante el código de *software* para descubrir vulnerabilidades que los delincuentes podrían utilizar para tener acceso, aunque los humanos aún tienen que aprobar cualquier cambio. Hace poco un bot de IA que fue creado por investigadores de Stanford superó a algunos expertos humanos en la búsqueda de fallas de seguridad en una red.

No obstante, incluso la IA no podrá impedir todas las infracciones, que es la razón por la que las organizaciones tienen que enfocarse en crear redes resistentes que puedan seguir funcionando incluso cuando estén bajo ataque, precisa Hultquist, el analista de Google.

Para los trabajadores de oficina que manejan en forma rutinaria correos electrónicos, envían y reciben documentos, o cualquier persona en casa en su computador personal, una dosis saludable de escepticismo y algunos buenos hábitos en línea pueden ser aún la mejor defensa. No abran archivos adjuntos sospechosos a menos que haya verificado el remitente, independientemente del correo electrónico original. La autenticación multifactor es muy eficaz, según expertos. Y si recibiera un correo de voz, correo electrónico o video de su jefe o un familiar en que le pide que le transfiera dinero o pague una cuenta de la que usted no tenía conocimiento, verifique con ellos primero para asegurarse de que es verdadero.

"La IA generativa logra hacer falsificaciones tan convincentes que el escepticismo es su mejor defensa", señala Marwick, la experta en datos del consumidor.

Artículo traducido por "El Mercurio".