Fuente:
Tomado de Juventud Rebelde

Era de esperar, aunque como reza el refrán, guerra avisada… no mata a soldado. ¿O sí? Con las mejoras constantes que muestra la inteligencia artificial a partir de los modelos de lenguaje extenso que han dado vida a chatbots como ChatGPT, de OpenAI, las conversaciones o el discurso humano es simulado con un realismo tremendo.

Y si bien esto tiene múltiples usos positivos, los cibercriminales ya han comenzado a emplearlos con fines negativos. Tal es el caso que ya algunos expertos consideran que nos adentramos en una nueva etapa de ataques sofisticados, los cuales serán muy difíciles de detectar en tanto las comunicaciones parecerán genuinamente humanas… o provenientes de una fuente confiable.

Jeremy Fuchs, investigador de la empresa de seguridad de correo electrónico Avanan, considera que lo que se conoce como el «correo electrónico comercial comprometido» (BEC, por sus siglas en inglés), un tipo de ataque mediante el cual se estafa a una empresa, ha entrado en su fase 3.0.

Con los años que ya tiene internet y su uso generalizado, los correos electrónicos han dejado de ser una fuente significativa de ataques de phishing, la técnica mediante la cual el usuario brinda información sensible porque cree que el mensaje arribó desde una fuente confiable, como un banco, su proveedor de bienes, un amigo, un familiar, entre otras. Aunque todavía se genera mucho spam, la mayoría de las personas emplean el sentido común para descartar las promesas de habernos ganado la lotería, un fabuloso viaje, o entregar información confidencial. Además, existen muchas herramientas automatizadas que detectan este tipo de correos, los clasifican como nocivos y los bloquean.

Sin embargo, han comenzado a aparecer estafas de phishing BEC que eluden estos sistemas de filtrado porque los correos provienen de direcciones legítimas, lo que significa que muchas de las defensas dependerán de que los empleados sean cautelosos y escépticos, y esto último, lo sabemos, muchas veces nunca se cumple en internet.

Los ataques BEC 3.0 se basan en la utilización de los servicios legítimos de un sitio web, por ejemplo, PayPal, para
compartir un archivo. Un pirata informático crea una cuenta y una factura, tal vez con un número de teléfono para iniciar un fraude telefónico. Y ahí inicia la estafa.

En otro ejemplo de ataque, encontrado y compartido por Avanan en marzo, se utilizaron los comentarios que aparecen en Google Workspace para enviar redireccionamientos maliciosos.

Si bien el consejo típico incluye estar atentos a los dominios de correo electrónico falsificados, la evolución

de esta táctica utiliza dominios familiares y legítimos. Y el contenido de los correos electrónicos ha sido refinado mediante inteligencia artificial, por lo que parecen bien creíbles.

En 2022 el Centro de Quejas de Delitos en Internet del FBI recibió 21 832 denuncias de BEC con pérdidas de más de 2 700 millones de dólares, según el boletín electrónico IT Brew.

Fuchs dijo a esa misma publicación que él y su equipo en Avanan atestiguaron más de 20 000 ataques de estilo BEC 3.0 en los primeros dos meses de 2023.

La era de la evolución

Los atacantes crean correos electrónicos de phishing utilizando el aprendizaje automático. En los foros de la deep web (la web profunda, el espacio de internet donde proliferan los hackers y otros criminales), se promociona la venta de estos servicios, según un artículo publicado en Demakis Technologies. Mencionan emplear el aprendizaje automático para producir correos electrónicos de phishing más efectivos. Operan mediante la creación de personalidades falsas para su uso en esfuerzos de estafa.

Los piratas informáticos pueden utilizar el aprendizaje automático para alterar de forma creativa los correos electrónicos de phishing para que no aparezcan en las listas de correo electrónico prohibidos y estén optimizados para fomentar la participación y los clics. Y la cuestión va más allá del texto: producen imágenes realistas, personajes de redes sociales y otros contenidos, todo con inteligencia artificial, en aras de lograr la mayor legitimidad posible.

Además, los delincuentes se apoyan en la IA y el aprendizaje automático para mejorar sus habilidades de hackeo de contraseñas. Es evidente que los motores de adivinación de contraseñas ahora tienen técnicas más sofisticadas basadas en la frecuencia y las tasas de éxito de los intentos de piratería criminal, explica una publicación en la web de la empresa Demakis Technologies.

Asimismo, utilizan el aprendizaje automático para identificar medidas de seguridad y adivinar mejores contraseñas con menos intentos, lo que aumenta su probabilidad de éxito.

Están, por otro lado, las herramientas de deepfake, capaces de producir videos o audios difíciles de distinguir del habla humana real. Es la forma más aterradora en que se emplean estas tecnologías por parte de los ciberdelincuentes.

Recientemente han salido a la luz algunos casos de alto perfil en los que se usaron audios falsos para estafar, y ello costó a las empresas cientos de millones de dólares. Desde 2016 las estafas por correo electrónico de las empresas han causado pérdidas por más de 43 000 millones de dólares, según Demakis Technologies.

Otro aspecto a tomar en cuenta es la ingeniería social para engañar y convencer a las víctimas de que revelen detalles confidenciales o realicen una acción específica, como enviar dinero al extranjero o abrir un archivo infectado.

Cómo resguardarse

Desde empresas como PayPal advierten que para protegerse de estos fraudes lo primero es estar siempre atentos a cualquier actividad sospechosa, así provenga el correo electrónico de una fuente supuestamente confiable.

Es bueno tomarse un tiempo para dilucidar preguntas: ¿De quién proviene el mensaje? ¿Ya se pagó la factura? ¿Se requiere un inicio de sesión? ¿Debo ingresar información personal o empresarial?

No está de más refrendar cualquier información adicional, ya sea con una llamada telefónica o preguntando en persona al supuesto remitente del mensaje si efectivamente pidió datos sensibles.

En todo caso el panorama no parece que vaya a hacerse más fácil, todo lo contrario. Los expertos predicen que el uso masivo de la IA generativa, como ChatGPT, proporciona vías para que los atacantes elaboren y

despliegen tácticas más sofisticadas a una mayor velocidad.

La empresa Gen Digital considera que tales estafas seguirán aumentando porque son fáciles de generar para personas con poca habilidad técnica.

Y basta con surfear la web para encontrar, casi a diario, nuevas herramientas para emplear la inteligencia artificial más allá de la generación de texto, lo que abre un abanico de posibilidades para los estafadores en línea. Se habla, incluso, de estafas en tiempo real, con la generación de videos o audios a partir de software de IA, algo que, aunque todavía es caro, solo se abaratará con el tiempo, como sucede con la mayoría de estas tecnologías.

Las empresas y organismos de seguridad tienen por delante un reto mayúsculo. El pulso recién comienza.