OpenAI informó que desmanteló una operación de fraude con sede en Camboya que utilizaba su chatbot de inteligencia artificial generativa (IA), ChatGPT, para facilitar una amplia variedad de estafas relacionadas con inversiones, romances, juegos de azar y la suplantación de autoridades policiales.

Como parte de esta acción, la compañía bloqueó una red coordinada de cuentas de ChatGPT que, presuntamente, se originaba en el sudeste asiático y operaba desde la ciudad de Poipet, una zona conocida por sus vínculos históricos con complejos dedicados a estafas y redes de trata de personas.

Según OpenAI, este conjunto de cuentas empleó sus modelos de IA para crear y gestionar identidades falsas en línea, redactar y traducir mensajes dirigidos a las víctimas, producir contenido promocional para actividades fraudulentas y apoyar diversas tareas operativas de la organización.

Entre el material generado se encontraban anuncios en redes sociales destinados a reclutar personal para puestos de «chatter» en Poipet, orientados principalmente a usuarios de Bangladesh e India. Estas ofertas prometían un salario base de 800 dólares, un bono adicional de 100 dólares por asistencia perfecta, además de boletos de avión, alojamiento y alimentación gratuitos, visas camboyanas con vigencia de un año y permisos de trabajo.

Asimismo, una parte de estas cuentas utilizó la herramienta de IA para realizar labores administrativas, como la redacción de comunicados internos, la traducción de mensajes entre empleados y el registro de deudas del personal, descuentos salariales, multas, pagos de préstamos, excedentes en el tiempo de permanencia con visa, permisos laborales, estatus migratorio e incentivos para el reclutamiento.

La empresa desarrolladora de IA señaló que la investigación se llevó a cabo en colaboración con WhatsApp, plataforma propiedad de Meta.

«Esta operación demuestra una realidad importante sobre las redes modernas de estafadores: los grupos delictivos organizados rara vez se limitan a un solo tipo de fraude»,

indicó OpenAI. «En su lugar, aprovechan cualquier narrativa, identidad o táctica que consideren más efectiva para engañar a sus víctimas.»

Los actores responsables de esta red ejecutaban múltiples modalidades de fraude de manera simultánea, combinando técnicas de distintos esquemas con el objetivo de incrementar la probabilidad de éxito de sus actividades ilícitas.

Además de utilizar perfiles falsos de citas para generar confianza antes de dirigir a las víctimas hacia inversiones fraudulentas relacionadas con criptomonedas y operaciones de oro al contado, los delincuentes mantenían prolongadas conversaciones románticas mediante identidades sintéticas o se hacían pasar por representantes de plataformas de apuestas en línea que ofrecían supuestos bonos y premios inexistentes.

Otros integrantes de la red se hacían pasar por organismos de seguridad pública para generar una sensación de urgencia, informando falsamente a las víctimas que debían pagar multas por presuntos delitos graves.

Para ejecutar estas estafas, la organización creó y administró perfiles ficticios en aplicaciones de citas, supuestos expertos en inversiones y falsas identidades de agentes policiales. Además, generó imágenes de documentos falsificados, incluidos pasaportes, notificaciones legales, comprobantes de compra de acciones e interfaces fraudulentas de plataformas de apuestas.

De acuerdo con OpenAI, la cadena de ataque sigue una metodología de tres etapas denominada *ping-zing-sting*. En primer lugar, los estafadores establecen contacto inicial mediante plataformas de mensajería como WhatsApp y Telegram; posteriormente, desarrollan una relación de confianza con las víctimas y, finalmente, las convencen de realizar depósitos, pagar cuotas de activación o cubrir multas inexistentes. Como respaldo aparente de las transacciones, proporcionan capturas de pantalla falsas de transferencias bancarias o información de cuentas.

Algunas cuentas pertenecientes a esta red también generaban contenido relacionado con la trata de personas y el trabajo forzado, prácticas asociadas con organizaciones criminales del este asiático. Estos grupos suelen atraer a personas mediante falsas ofertas de empleo altamente remuneradas, para posteriormente confiscarles sus pasaportes y obligarlas a trabajar en condiciones equiparables a la esclavitud.

«La magnitud total de las pérdidas económicas ocasionadas por esta red aún es desconocida; sin embargo, con base en las propias comunicaciones de los estafadores, es posible que la operación haya interactuado con cientos de víctimas a través de distintos tipos de fraude», señaló OpenAI. «Las conversaciones de los usuarios hacían referencia a pérdidas individuales de miles de dólares, aunque no hemos podido verificar de manera independiente dichas afirmaciones.»

Este caso evidencia cómo los actores maliciosos están incorporando capacidades ofensivas potenciadas por inteligencia artificial, lo que les permite incrementar significativamente la velocidad y el alcance de sus campañas. Al mismo tiempo, incluso los modelos de IA más avanzados han demostrado ser capaces de interactuar con sistemas e individuos reales durante evaluaciones de ciberseguridad del tipo *Capture-the-Flag (CTF)*.