Las aplicaciones que generan deepfake explotan y permiten atracos corporativos multimillonarios

Los deepfakes se están volviendo rápidamente más realistas y el acceso a ellos más democrático, lo que permite que incluso los atacantes comunes cometan fraudes importantes. ¿Cuál es la forma más eficaz de contraatacar?

Imagen de Nate Nelson, escritor colaborador

Nate Nelson, escritor colaborador

5 de febrero de 2024

Lectura de 4 minutos

Rostro superpuesto a un cuerpo, que ilustra deepfakes

FUENTE: MIKE A TRAVÉS DE ADOBE STOCK

El software de creación de deepfake está proliferando en la Dark Web, lo que permite a los estafadores llevar a cabo fraudes financieros asistidos por inteligencia artificial (IA) con una creatividad y un alcance nunca antes vistos.

Consideremos lo que sucedió hace unas semanas, cuando un empleado del departamento de finanzas de una corporación multinacional con sede en Hong Kong recibió un mensaje. Era el director financiero de su empresa con sede en el Reino Unido, que le pedía que llevara a cabo una transacción. Según informa The South China Morning Post, el empleado tuvo un "momento de duda" inicial sobre la solicitud, hasta que algo le hizo cambiar de opinión.

Poco después del mensaje inicial, el empleado realizó una videoconferencia con ese director financiero, junto con una lista de otros colegas. Todos parecían y sonaban como las personas que conocía. Se le pidió que hiciera una breve introducción al grupo, luego se le dieron instrucciones y la reunión terminó abruptamente.

En la semana siguiente, el empleado continuó discutiendo las transacciones con sus "colegas" a través de mensajería instantánea, correo electrónico y llamadas individuales. Cuando se reveló el engaño, ya había realizado 15 transacciones por un total de 25,5 millones de dólares (unos 200 millones de dólares de Hong Kong).

El floreciente mercado del software deepfake

Los deepfakes (también los buenos) existen desde hace algún tiempo. "Esto es algo que la gente ha podido hacer durante varios años. Creo que fue en diciembre de 2020 cuando dije por primera vez, después de pasar una década mirando falsificaciones, que ya no podía notar la diferencia", dice Dominic Forrest. CTO de iProov.

Lo que ha cambiado es el hecho de que están más disponibles para un público más amplio, con una barrera de entrada más baja.

En la década de 2010, explica, "era un trabajo altamente cualificado, que requería mucho procesamiento y era muy difícil de realizar. Hoy en día, hay muchas herramientas sencillas en el mercado que la gente puede descargar de forma gratuita o por unas pocas decenas de dólares, que crean deepfakes de calidad de personas solo a partir de una única imagen de referencia de un perfil de LinkedIn, una página de Facebook o Twitter, dondequiera que esté".

El intercambio de caras, por ejemplo, se ha vuelto absolutamente común. Para un informe que se publicará el miércoles, iProov ha rastreado más de 100 herramientas distintas en el mercado hoy en día diseñadas para crear intercambios de caras simples.

También existen ofertas más avanzadas, como OnlyFake, un servicio de la Dark Web que puede producir una identificación falsa realista en un instante, o muchas de ellas a escala, por sólo 15 dólares cada una. Una publicación reciente en su cuenta de Telegram anuncia el cambio radical y afirma que "la era de renderizar documentos usando Photoshop está llegando a su fin".

Estos mismos avances en calidad y accesibilidad también han permitido el florecimiento de productos ultrafalsos. Las huelgas de Hollywood en 2023 fueron impulsadas en parte por preocupaciones sobre la aplicación de esta tecnología a las películas y la televisión, que podría hacer que los extras queden obsoletos, y el gigante multimedia chino Tencent ofrece ahora un servicio comercial de deepfake capaz de crear falsificaciones humanas realistas y de alta definición utilizando solo tres minutos de video de acción en vivo y 100 oraciones habladas como material fuente.

Las soluciones fáciles para la detección de deepfake y las difíciles

Gran parte del discurso en torno a la seguridad de los deepfake se centra en identificar idiosincrasias en su producto final: las imperfecciones en una imagen falsa, la falta de resonancia que podría revelar una voz generada por IA y otras deficiencias técnicas que podría tener un software humano o anti-deepfake. poder marcar como sospechoso.

Sin embargo, debido a que la tecnología está mejorando tan rápidamente, esto se vuelve cada vez más difícil de lograr.

"Mucha gente sugiere cosas como: haz que la gente se quite y se vuelva a poner las gafas y podrás detectar si se trata de un deepfake. Eso ya no es cierto. Como ser humano, no puedes notar la diferencia. ", dice Forrest.

Intentar superar el software puede ser una estrategia que valga la pena, afirma Kevin Vreeland, director general de Veridas en Norteamérica. Pero ofrece una alternativa aún más simple y confiable para lidiar con los deepfakes en un nivel más fundamental: en lugar de preguntar constantemente si todo es real, las empresas pueden concentrarse en evitar que los productos sintéticos lleguen a los empleados en primer lugar.

"Tal vez están llamando desde un número de teléfono desconocido", dice sobre los atacantes deepfake, "o llamando desde una ubicación desconocida. Incluso puedes exigir a las personas que activen la geolocalización en sus dispositivos para realizar estas grandes transferencias de dinero. Una vez que Si empiezas a desglosar todo eso, verás que los datos que vienen con

No es legítimo y ahí es donde hay que empezar a cuestionar y denunciar".

Hasta que la tecnología de detección se ponga al día, son estos metadatos más básicos los que facilitan la elección. Porque si bien sus deepfakes son notablemente realistas, señala Vreeland, "estos atacantes de deepfake no pueden cubrir todas sus bases".