Tecnología Deepfake:
¿Qué es y cómo detectar el contenido malicioso?

CIUDAD DE MÉXICO. 17 de abril de 2023.- Como si se tratara del villano de un filme cinematográfico de ciencia ficción, muchas de las amenazas utilizan soluciones tecnológicas que de origen son benignas, pero su uso se desvirtúa al grado que los cibercriminales las adoptan como herramientas comunes para propagar ataques.

Un ejemplo de esto es la tecnología *Deepfake*, que fue creada con fines legítimos como la creación de avatares en videojuegos o en diferentes plataformas *online*, así como la impartición de clases remotas; pero que también ha sido utilizada para engañar a la gente con fines de extorsión, entre otros.

"Se trata de videos y/o imágenes que imitan la apariencia y la voz de una persona, como son los filtros de realidad aumentada en diversas plataformas digitales, así como los avatares que utilizan las empresas en capacitaciones y/o tutoriales. En ocasiones pueden llegar a ser tan precisos, que los cibercriminales los han adoptado para convertirlos en una herramienta de engaño", explica Javier Bernardo, Head of Strikers de Strike.

Strike aclara que el *deepfake* es generado por Inteligencia Artificial (IA) y funciona bajo el aprendizaje profundo, que le ayuda a convertirse en una solución cada vez más humana e intuitiva en su interacción. Desde el año pasado se han presentado casos de *deepfakes* tan convincentes al momento de interactuar, que pueden engañar a las personas y a los propios algoritmos.

Prueba de que se están volviendo cada vez más difíciles de distinguir, es la encuesta realizada por iProov el año pasado que indica que a nivel global el 71% no sabe que son los *Deepfake* y no podrían distinguirlos. Añade que México, junto con el Reino Unido, es el país más familiarizado con este concepto, con un 40% de los encuestados.

Dicha fuente cita algunos ejemplos populares el año pasado, como diversos videos de Marck Zuckerberg, propietario de Meta, realizando anuncios falsos; y un video de la Reina Isabel II dando un mensaje navideño falso, en diciembre de 2020.

Un reporte de Kyriba indica que el año pasado se incrementó el uso de esta tecnología en las estafas dirigidas a correos empresariales, mediante audios enviados a los colaboradores en los que se les hace creer que quien habla es algún directivo, obteniendo así accesos a datos financieros y credenciales sensibles.

Los *deepfakes* también podrían utilizarse para propagar desinformación. Por ejemplo, un video con la voz o la imagen de un directivo de una compañía expresando un mensaje de odio. Este escenario lo destaca Europol como uno de los fines de mayor utilización este año.

Al respecto, Strike recomienda crear campañas de concientización sobre los *Deepfake*, y cómo es que la IA puede ser utilizada con el objetivo de engañar. Esto mantendrá alerta a la plantilla y ayudará a detectar de mejor forma cualquier contenido que, en otro escenario, podría convertirse en una amenaza.

Por otra parte, y más allá de la ingeniería social, Strike recomienda a los equipos de ciberseguridad que generen estrategias de detección de contenidos tanto manuales como automatizados.

Por una parte, es posible hacerlo manualmente mediante la detección de inconsistencias que, dependiendo el nivel de sofisticación del cibercriminal, pueden ser notorias en primera instancia. Imágenes mal enfocadas alrededor del rostro; parpadeo inusual o falta del mismo; reflejo de la luz en los ojos; inconsistencias en detalles como el movimiento del cabello; y desde luego errores en cuanto al desenfoque del fondo o la imagen del *background*, son aspectos en los que deben estar pendientes.

La detección automatizada, por otro lado, implica la implementación de soluciones como el escaneo de los sistemas en los que se puede detectar con precisión el origen de los archivos en video, imagen y audio que ingresan al sistema para encontrar, desde esa perspectiva, las anomalías que los hacen relucir como contenido apócrifo. En esta tarea, los *hackers* éticos pueden ser de gran ayuda ya que sus conocimientos proveen a la compañía una visión distinta sobre cómo abordar el problema.

En conclusión, la tecnología Deepfake, bien empleada, puede ayudar a las compañías a obtener una serie de beneficios en materia de optimización del tiempo y los recursos, gracias a la Inteligencia Artificial. Pero al mismo tiempo es importante saber que nunca se debe descartar la posibilidad de ser vulnerado mediante un video o un audio apócrifo, por lo que es clave estar alerta para anticiparse a cualquier potencial estafador.

-o0o-

Sobre Strike

Strike es una plataforma de ciberseguridad en Latinoamérica. Su principal misión es ayudar a que las compañías estén protegidas a través de la detección y resolución de vulnerabilidades en sus sistemas. Esto se realiza a través de tests de penetración - o *pentests* - llevados a cabo por su red global de hackers éticos, conocidos como “Strikers”, una comunidad global que reúne a los mejores expertos de ciberseguridad con reconocimientos y certificaciones internacionales. Su objetivo es impulsar una cultura de ciberseguridad de calidad y accesible, en la que la misma sea parte del ciclo de vida de las empresas y no algo ocasional o independiente. Más información en: https://strike.sh/

Síguenos en nuestras redes sociales:
Instagram - @strikesecurity
Twitter - @strike_secure
LinkedIn - Strike

Contacto para prensa México
another
Ahtziri Rangel | PR Expert
+ 52 1 55 1395 6970
ahtziri.rangel@another.co