Aprende una nueva habilidad de supervivencia: Detectando deepfakes

¿Qué son los deepfakes?

La palabra "deepfake" es una combinación de "deep learning" (aprendizaje profundo) y "fake" (falso). Deepfakes son imágenes, videos o grabaciones de audio falsificados. A veces, las personas que aparecen en ellos son identidades falsas generadas por computadora que se ven y suenan como si pudieran ser personas reales. Otras veces las personas son reales, pero sus imágenes y voces son manipulados para hacer y decir cosas que no hicieron ni dijeron. Por ejemplo, un video falso podría usarse para recrear a una celebridad o un político diciendo algo que nunca dijeron. Usando estas falsificaciones realistas, los atacantes pueden generar una realidad alternativa en la que no siempre puedes confiar en tus ojos y oídos.

Algunos deepfakes tienen propósitos legítimos, como películas que regresan a la vida a actores fallecidos para recrear algún personaje famoso. Pero los atacantes cibernéticos están comenzando a aprovechar el potencial de los deepfakes. Los despliegan para engañar a tus sentidos, para que puedan robar tu dinero, acosar a la gente, manipular a los votantes o las opiniones políticas, o crear noticias falsas. En algunos casos, incluso han creado empresas fraudulentas formadas por empleados falsos. Debes ser aún más cuidadoso con lo que crees cuando lees noticias o redes sociales a la luz de estos ataques.

El FBI advierte que en el futuro los deepfakes tendrán "un impacto más severo y generalizado debido al nivel de sofisticación de los medios sintéticos utilizados". Aprende a detectar los signos de un deepfake para protegerte de estas simulaciones altamente creíbles. Cada forma de deepfake (imagen fija, video y audio) tiene su propio conjunto de fallas que pueden delatarlo.

Imágenes estáticas

El deepfake que puedes ver con más frecuencia es la típica imagen de perfil falsa de redes sociales. La imagen a continuación es un ejemplo de un deepfake del sitio web thispersondoesnotexist.com. En la imagen hay cinco pistas diferentes que te indican que podría ser un deepfake. Notarás que estas pistas no son fáciles de detectar y pueden ser difíciles de identificar:

1. Fondo: Este a menudo está borroso o torcido, y puede tener una iluminación inconsistente, como sombras pronunciadas que apuntan en diferentes direcciones.
2. Lentes: Mira de cerca la conexión entre los lentes y los marcos cerca de la sien. Los deepfakes a menudo tienen conexiones que no coinciden con tamaños o formas ligeramente diferentes.
3. Ojos: Los deepfakes que se usan actualmente para las fotos de perfil falsas parecen tener el brillo en los ojos en el mismo lugar, lo que resulta en lo que algunos llaman la "mirada del deepfake".
4. Joyas: Los aretes pueden no tener forma o estar extrañamente unidos. Los collares pueden parecer incrustados en la piel.
5. Collares y hombros: Los hombros pueden estar deformados o no al mismo nivel. Los collares pueden ser diferentes en cada lado.

Video

Investigadores del Instituto de Tecnología de Massachusetts (MIT) desarrollaron una lista de preguntas para ayudarte a determinar si un video es real, y señalaron que los deepfakes a menudo no pueden "representar completamente la física natural" de una escena o iluminación.

1. Mejillas y frente: ¿La piel parece demasiado suave o demasiado arrugada? ¿La edad de la piel es similar a la edad del cabello y los ojos?
2. Ojos y cejas: ¿Aparecen sombras en los lugares que esperarías?
3. Lentes: ¿Hay algún destello? ¿Demasiado brillo? ¿Cambia el ángulo del reflejo cuando la persona se mueve?
4. Vello facial: ¿El vello facial parece real? Los deepfakes pueden agregar o eliminar bigote, patillas o barba.
5. Lunares faciales: ¿El lunar parece real?
6. Parpadeo: ¿La persona parpadea lo suficiente o demasiado?
7. Tamaño y color de los labios: ¿El tamaño y el color coinciden con el resto de la cara de la persona?

Audio/voz

Los investigadores dicen que tecnologías como los espectrogramas pueden mostrar cuándo las grabaciones de voz son falsas. Pero la mayoría de nosotros no tenemos el lujo de contar con un analizador de voz cuando llama un atacante. Escuchas una forma de hablar monótona, un tono o una emoción extraña y la falta de ruido de fondo. Las falsificaciones de voz pueden ser difíciles de detectar. Si recibes una llamada extraña de una organización legítima, puedes verificar si la llamada es real colgando primero y luego devolviendo la llamada a la organización. Asegúrate de usar un número de teléfono confiable, como un número de teléfono que ya tengas en tu lista de contactos, un número de teléfono impreso en una factura o estado de cuenta de la organización, o el número de teléfono del sitio web oficial de la organización.

Conclusión

Ten en cuenta que los atacantes están usando deepfakes activamente. Pueden crear cuentas falsas en redes sociales para conectarse o crear videos falsos para influir en la opinión pública. Algunos incluso están vendiendo sus servicios en la dark web para que otros atacantes puedan hacer lo mismo. No esperamos que te conviertas en un experto en deepfake, pero si te armas con los conceptos básicos para identificar dichas falsificaciones serás mucho mejor defendiéndote. Si sospechas que has detectado un deepfake, repórtalo al sitio web o fuente que aloja el contenido.

Editor invitado

Kerry Tomlinson (@KerryTNews) es reportera de cibernoticias en Ampere News y certificada como SANS Security Awareness Professional. Su misión es traducir lo que está sucediendo en el mundo digital para personas de todos los niveles de conocimiento con noticias convincentes y presentaciones perspicaces.

Recursos

Ingeniería social: https://www.sans.org/newsletters/ouch/social-engineering-attacks/
¿Puedes detectar lo falso? (Ampere News): https://www.amperesec.com/news/can-you-spot-the-fake
Prueba de detección de deepfakes del MIT: https://detectfakes.media.mit.edu/
Reconoce las deepfakes: https://www.spotdeepfakes.org/en-US

Traducido para la comunidad por: Célica Martínez Aponte and Iván Galindo, UNAM-CERT: