Una nueva frontera para el ciberdelito: clonación de voz y audio de IA

08/12/2020

La tecnología deepfake es una tecnología basada en la Inteligencia Artificial, pero mal intencionada supone algunos peligros entre la sociedad, como es engañar y crear confusión. La estafa de audio deepfake es sin duda una de las aplicaciones más sorprendentes de esta clase de tecnología. Incluso, se ejecuta de manera tan convincente que hasta un Gerente de Administración, víctima del ciberataque, ha declarado haber reconocido la voz de su jefe.
Dejando a un lado la tecnología sofisticada, el proceso detrás de la construcción de deepfakes de audio es muy sencillo. Los piratas informáticos, a través de un software espía y dispositivos que permiten recopilar varias horas de grabaciones de la víctima hablando, han ajustado la tecnología de aprendizaje automático para clonar la voz de un individuo. A mayor cantidad de datos recopilados, y mejor la calidad de las grabaciones, más precisa y potencialmente dañina será la clonación de voz en la práctica.

La IA utiliza lo que se conoce como redes generativas adversarias (GAN), unos sistemas que compiten continuamente entre sí, a través de los cuales uno crea una falsificación y el otro intenta identificar sus defectos. Con cada nuevo intento, la IA puede mejorarse exponencialmente hasta lograr una mímica confiable y, a menudo, precisa y exitosa tras analizar tan solo veinte minutos de grabación.

Lamentablemente, estas grabaciones son muy fáciles de recopilar, especialmente en épocas de COVID-19, donde el contacto físico se ha perdido y las horas de comunicación digital son permanentes. Los discursos online compartidos a

través de las redes sociales, las llamadas telefónicas, las entrevistas, reuniones y videoconferencias, son relativamente fáciles de acceder y una fuente inagotable de materia prima para procesar. Esto preocupa a muchos empleados, funcionarios y ejecutivos, sobre todo de grandes empresas, ya que los piratas informáticos de audio deepfake pueden eludir la más robusta de las protecciones de la red informática, pueden protegerse contra la mayoría de malware y virus sofisticados, y actualizar constantemente las VPNs para detectar nuevas amenazas y tipos de virus.

Sin embargo, como usuarios agregamos involuntariamente una enorme cantidad de información en redes sociales por error o por impericia. Si bien los ciberdelincuentes pueden penetrar nuestras defensas, el combo sumado a nuestros errores, es por lo menos peligroso.

Una de las alternativas de protección es la aplicación de algoritmos complejos y de gran alcance que tienen la capacidad de aprender patrones y peculiaridades del habla humana para hacer usados en la detección de pistas de audio deepfake. Otra alternativa, más práctica y efectiva de identificar una estafa de deepfake, es simplemente devolver la llamada.

La mayoría de este tipo de estafas se llevan a cabo con el uso de una cuenta VOIP de grabación, configurada para contactar a los objetivos en nombre de los piratas informáticos. Al devolver la llamada, las víctimas pueden averiguar de inmediato si está detrás una persona real o no.

En la actualidad, la tecnología no está lo suficientemente extendida como para que las estafas de audio sean una preocupación de gran alcance. No obstante, la IA evoluciona a un ritmo vertiginoso, y la tecnología que hace posible la simulación profunda sea cada vez más accesible y fácil de usar.

Esencialmente, los sistemas de seguridad del futuro cercano

serán imitaciones avanzadas de las mismas herramientas de inteligencia artificial que los piratas informáticos maliciosos están utilizando en sus intentos de defraudar a sus víctimas. El mejor consejo es permanecer alerta y preparado, ya que las estafas de audio deepfake podrían convertirse en el próximo gran problema.

(*) CEO de BTR Consulting, Especialista en ciberseguridad, riesgo tecnológico y de negocios

Fuente: Ámbito