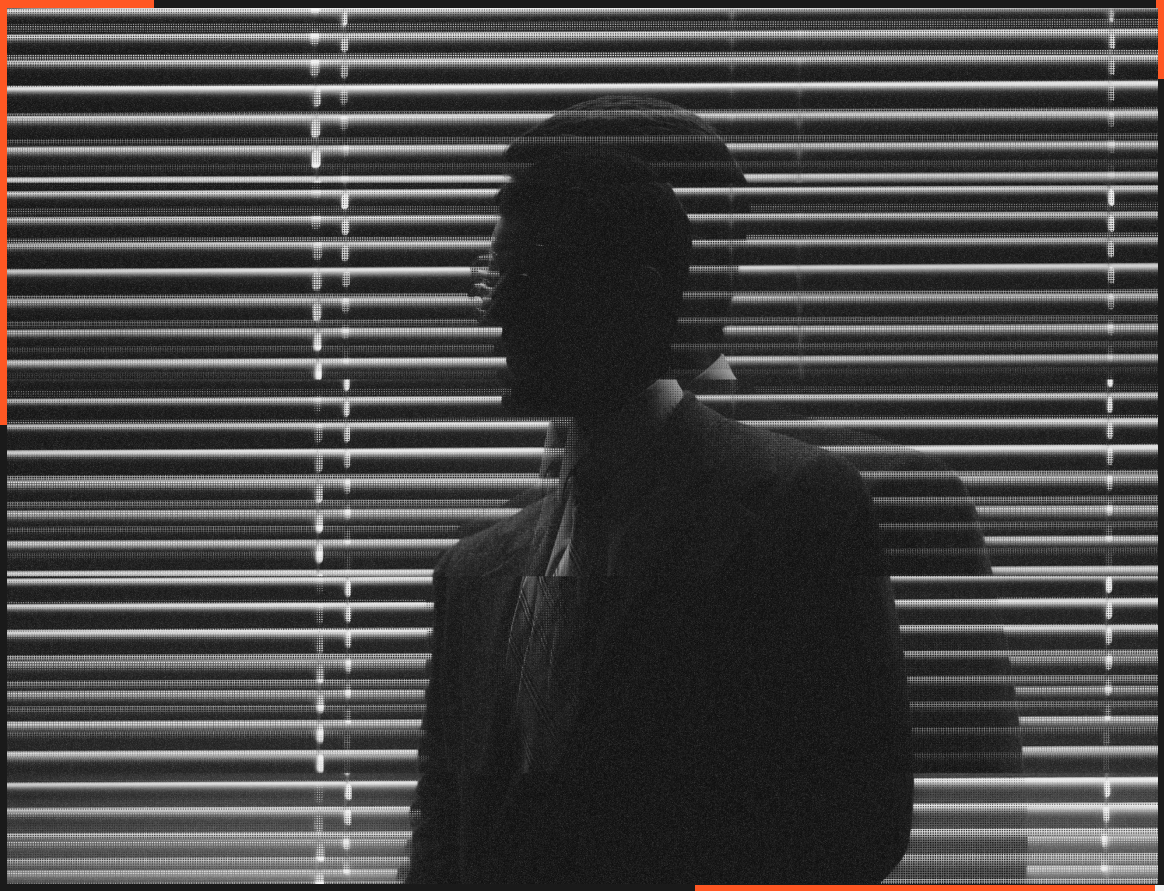


E-book

Deepfakes en campañas de fraude corporativo

Cómo la simulación de ejecutivos se ha convertido en una amenaza directa para la reputación y la seguridad de las empresas

AXUR

Resumen Ejecutivo

El avance de las tecnologías de inteligencia artificial generativa —especialmente los deepfakes— ha puesto de manifiesto una amenaza invisible pero poderosa: la manipulación de la identidad de líderes y figuras de autoridad como arma de fraude.

Ejecutivos de alto nivel se han convertido en objetivos estratégicos en campañas que exploran una nueva dimensión de vulnerabilidad: su imagen pública.

De videollamadas con voz clonada a perfiles falsos en redes sociales, los estafadores están superando barreras tecnológicas y psicológicas para engañar a colaboradores, mover grandes sumas de dinero y comprometer la reputación de las organizaciones.

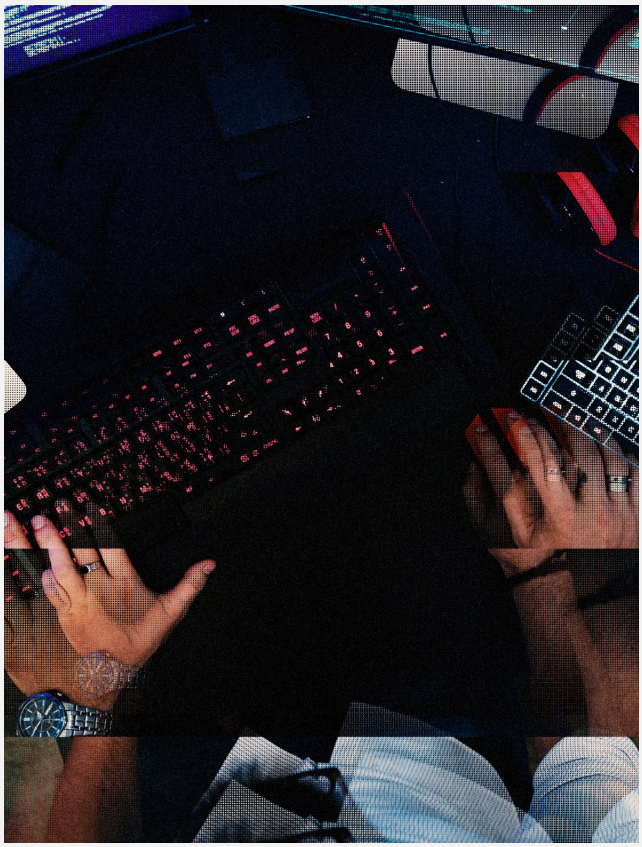
En este eBook, exploramos cómo se estructuran estas campañas, por qué los ejecutivos son el blanco principal, y cómo las empresas pueden protegerse.

Basándonos en nuestro framework propietario Fraud Neuron, decodificamos los patrones y tácticas detrás de estas operaciones y mostramos cómo la inteligencia artificial puede ser utilizada tanto para el ataque como para la defensa.

Deepfakes: la ilusión perfecta al servicio del fraude

Los deepfakes dejaron de ser curiosidades tecnológicas o amenazas limitadas a celebridades. Hoy en día, alimentan fraudes corporativos a gran escala, utilizando la imagen y la voz de ejecutivos para convencer a colaboradores de tomar decisiones críticas, como autorizar transferencias bancarias, abrir accesos a sistemas o divulgar información estratégica.

Con una sola imagen pública y unos pocos segundos de audio, los criminales logran generar avatares realistas que hablan, gesticulan y expresan emociones de manera convincente. La naturaleza de estas campañas es estratégica: se basan en ingeniería social, en la autoridad de la figura ejecutiva y en situaciones de urgencia para comprometer la racionalidad de la víctima.



Por qué los ejecutivos son los nuevos objetivos preferenciales

No es coincidencia que CFOs, directores regionales y CEOs estén en el centro de estas fraudes. Reúnen tres características que los hacen valiosos para operadores maliciosos:

1. Alto nivel de exposición pública: entrevistas, videos institucionales, eventos online y redes sociales proporcionan abundante material para el entrenamiento de modelos de deepfake.
2. Autoridad incuestionable: una orden proveniente del “CFO” por video o audio rara vez es cuestionada por los colaboradores, especialmente bajo presión.
3. Acceso a recursos críticos: los ejecutivos suelen poseer o influir en decisiones financieras y operativas sensibles.

Con estas variables combinadas, campañas bien orquestadas logran sortear controles técnicos, explotando una vulnerabilidad esencialmente humana: el respeto a la jerarquía. Además de fraudes financieros directos, hay indicios del uso de estas técnicas para obtener acceso inicial en entornos corporativos a través de social engineering, o ingeniería social dirigida.

Anatomía de un deepfake: más allá de la ilusión digital

Para entender la complejidad e impacto de este fenómeno, utilizamos aquí la estructura táctica de clasificación basada en Fraud Neuron, framework propietario de Axur desarrollado para describir, categorizar y monitorear fraudes digitales complejos.

A continuación, describimos cada una de las fases típicas de una campaña con deepfakes, según lo observado en incidentes reales.

→ Identificación de objetivos (T1000) – Target Identification

La campaña maliciosa realiza una identificación precisa de objetivos, priorizando tanto a individuos como a organizaciones de alto valor. Entre los **objetivos individuales (TQ1110)**, destacan **empleados (P1112)** con acceso a finanzas u operaciones críticas y **ejecutivos senior (P1113)**, como CFOs, que son utilizados como modelo para los deepfakes. En el **espectro corporativo (TQ1120)**, los actores apuntan a grandes corporaciones (P1122) con estructuras descentralizadas y cadenas de mando fragmentadas, lo que facilita la ingeniería social. También hay indicios de explotación de **instituciones financieras (P1123)**, especialmente mediante el uso de la identidad de ejecutivos de empresas del sector cripto.

→ Temas (T2000) – Themes

La narrativa de las fraudes se construye cuidadosamente en torno a temas financieros y de oportunidades profesionales. Los actores utilizan ofertas de inversión (P2113) como pretexto, por ejemplo, a través de deepfakes de ejecutivos de Binance promoviendo supuestas oportunidades en cripto. En paralelo, se exploran temas de empleo (TQ2200) mediante falsas ofertas laborales (P2211) difundidas en redes sociales, con videos deepfake de celebridades brasileñas promoviendo productos o servicios inexistentes. Estos temas tienen un alto atractivo y amplían el alcance de las campañas fraudulentas.

→ Reconocimiento (T3000) – Reconnaissance

La fase de reconocimiento implica una extensa recopilación de datos públicos (TQ3110). Los atacantes extraen información visual y vocal de fuentes abiertas (P3111), como entrevistas y presentaciones públicas, además de recolectar datos de redes sociales (P3113) para comprender lenguaje corporal, patrones de habla y contexto organizacional. Estos datos alimentan modelos de IA capaces de generar deepfakes convincentes. No hay indicios del uso de fuentes privadas ni de técnicas de recopilación técnica, pero la profundidad del contenido generado sugiere un enfoque meticuloso de open-source intelligence.

→ Recursos (T4000) – Resources

Los operadores implementan infraestructura dedicada para sus fraudes. Crean cuentas falsas en redes sociales (P4113) para simular la presencia de figuras públicas o ejecutivos. Se registran dominios (P4211) con nombres similares a las marcas corporativas para alojar contenido deepfake o páginas de phishing. La campaña utiliza infraestructura de servidores (P4212) para almacenar videos manipulados y realizar transmisiones en tiempo real. También se preparan billeteras de criptomonedas (P4214) para recibir fondos provenientes de las fraudes. La operación se respalda con sistemas de inteligencia artificial (P4311) para la generación de videos y softwares especializados (P4312) que realizan lip-sync, clonación de voz y manipulación de audio de alta fidelidad.

→ Conversión (T7000) – Conversion

La extracción de valor se realiza a través de transferencias bancarias fraudulentas (P7111) tras convencer a la víctima. En otros casos, el valor desviado se convierte y se mueve mediante transacciones en criptomonedas (P7113), dificultando su rastreo. Esta táctica también se aplica a fraudes con falsas promociones y ventas, donde el pago se dirige directamente a billeteras digitales controladas por los operadores de la campaña.

→ Simulación de identidad (T5000) – Identity Simulation

La simulación de identidad es el núcleo de la campaña. Los atacantes practican la personificación de empleados (P5111), especialmente de ejecutivos de alto nivel, en videollamadas y audios. También utilizan deepfakes de personalidades públicas (P5113) para generar videos promocionales falsos que circulan como anuncios pagados. Para ello, aplican técnicas de spoofing de canal (P5213), simulando llamadas por Zoom, Microsoft Teams y otras plataformas legítimas, creando una ilusión de autenticidad mediante la sincronización de imagen y voz. La combinación de estas técnicas resulta en el contacto directo con las víctimas en tiempo real o mediante videos altamente convincentes.

→ Ingeniería social (T6000) – Social Engineering

La campaña se fundamenta en ingeniería social sofisticada. Los operadores crean situaciones de emergencia (P6112) en las que el deepfake de un ejecutivo exige una acción inmediata, generalmente una transferencia bancaria o la liberación de accesos. Emplean presión por tiempo (P6121) para reducir la capacidad crítica de las víctimas, forzándolas a tomar decisiones rápidas. Al utilizar canales corporativos y la imagen de autoridad, los atacantes inducen presión social (P6123), haciendo que la víctima se sienta obligada a obedecer las órdenes del "superior". La manipulación psicológica se ve potenciada por el realismo de la imagen y la voz generadas por IA.

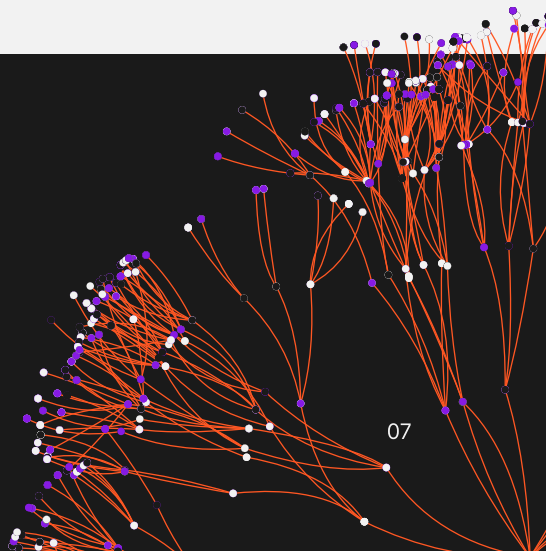
→ Impactos (T8000) – Impacts

Los impactos incluyen pérdidas financieras directas (P8111), como en el caso documentado de USD 25 millones desviados mediante una videoconferencia con un deepfake del CFO. También se observan daños a la reputación (P8211) cuando las marcas de las empresas son explotadas en campañas fraudulentas, como en los videos falsos difundidos con celebridades brasileñas. Finalmente, la exposición a estos engaños genera impacto psicológico (P8213) en los colaboradores engañados, que creyeron estar interactuando con sus superiores o con figuras públicas reales.

Fraud Neuron

Conozca más sobre Fraud Neuron y contribuya al mapeo de fraudes digitales:

github.com/axur/fraudneuron



DeepFake as a Service

El modelo Deepfake-as-a-Service consiste en la comercialización directa de tecnologías de medios sintéticos basadas en inteligencia artificial, enfocadas especialmente en la creación de videos, audios y avatares deepfake a demanda. Se trata de una estructura de negocio típica, en la que los proveedores ofrecen acceso simplificado y de bajo costo a tecnologías complejas para clientes sin conocimientos técnicos avanzados.

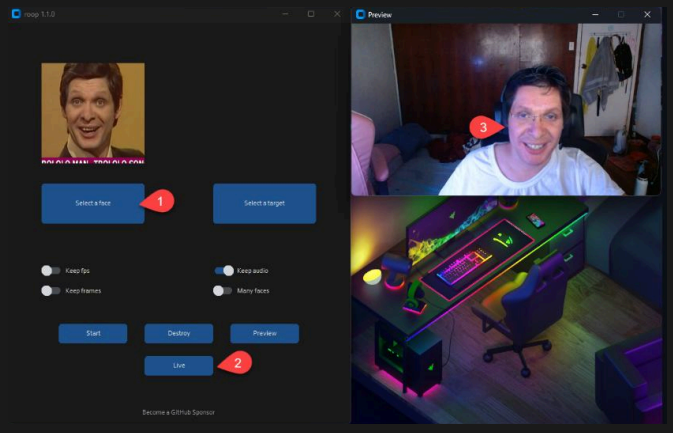
La característica esencial de este modelo es la disponibilidad de plataformas online donde los compradores solicitan servicios personalizados, tales como la creación de videos manipulados, llamadas en tiempo real con avatares deepfake o la replicación de voz para campañas específicas de fraude.

El mercado de DeepFake as a Service (DFaaS) ha crecido rápidamente en los últimos años, impulsado por la accesibilidad y simplicidad de las herramientas basadas en inteligencia artificial generativa. En este modelo de negocio, plataformas comerciales ofrecen tecnologías avanzadas que permiten la creación de contenidos manipulados realistas de manera sencilla, incluso para usuarios sin experiencia técnica significativa.

Al reducir drásticamente las barreras técnicas y financieras, estas herramientas no solo democratizaron el acceso a tecnologías sofisticadas, sino que también ampliaron significativamente el riesgo de uso malicioso de los medios sintéticos.

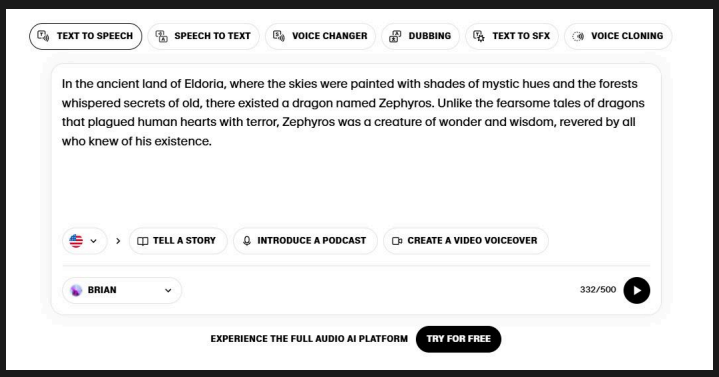
Una herramienta destacada en este segmento es Roop, una aplicación de código abierto que se popularizó por permitir reemplazar el rostro de cualquier persona en un video utilizando únicamente una imagen de referencia, sin necesidad de entrenar modelos ni recolectar grandes conjuntos de datos.

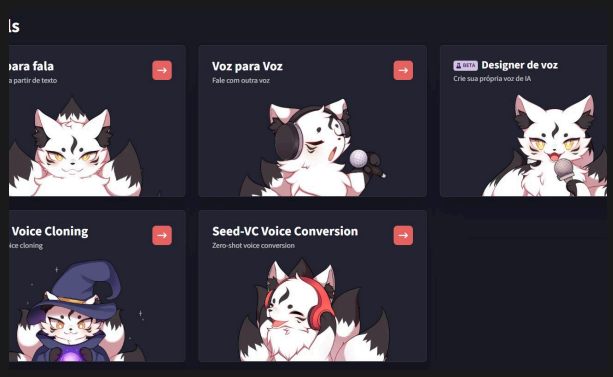
Roop utiliza redes neuronales preentrenadas para realizar sustituciones faciales casi instantáneas, produciendo resultados visuales convincentes incluso con una inversión técnica mínima (Roop, GitHub, 2023).



Además, existen plataformas que ofrecen clonación de voz y síntesis vocal avanzada, como ElevenLabs, que pueden ser explotadas de manera maliciosa. Con esta tecnología, los usuarios pueden crear audios realistas a partir de pequeños fragmentos vocales originales, replicando matices de emoción, ritmo y entonación con extrema fidelidad.

La popularización de esta herramienta ha sido mencionada como un factor decisivo en ataques exitosos de ingeniería social, como los casos documentados recientemente contra ejecutivos de grandes corporaciones, en los cuales la voz clonada fue utilizada para fraudes financieros significativos (CNN, 2024; Fortune, 2024).





Una herramienta similar, conocida como FakeYou, ofrece síntesis vocal simplificada a partir de texto escrito. FakeYou se diferencia por contar con un extenso catálogo de voces sintéticas preentrenadas, incluyendo personajes ficticios y celebridades, lo que facilita aún más el proceso de creación de contenido sintético.

La simplicidad de esta plataforma la ha vuelto popular no solo para usos legítimos, sino también como una herramienta rápida para campañas fraudulentas o de desinformación (FakeYou, 2023).

La herramienta conocida como Nudefusion utiliza inteligencia artificial para transformar fotos comunes en imágenes falsas con contenido explícito, en un proceso conocido como "deepnude". Este servicio genera especial preocupación debido a su potencial de abuso, invasión de privacidad y difamación. La herramienta se promociona de forma abierta, prometiendo generar resultados rápidos a partir de la simple carga de una imagen común, sin necesidad de entrenamiento previo ni habilidades técnicas específicas por parte del usuario.

Debido a su alto potencial de uso indebido, plataformas similares ya han sido objeto de investigaciones y acciones legales en diversos países, demostrando claramente el impacto social y ético de las tecnologías de deepfake cuando se aplican para la generación no consensuada de imágenes íntimas falsas.

Ataques que llegaron a los medios: casos reales y técnicas utilizadas

Escenario #1 Sabotaje corporativo

En este escenario, los criminales utilizan deepfakes para difundir desinformación, dañar la imagen de marcas, manipular la percepción del mercado o comprometer negociaciones de fusiones y adquisiciones (M&A). El sabotaje ocurre mediante la creación de videos y audios falsos que muestran a ejecutivos o empleados en situaciones comprometedoras o divulgando información incorrecta sobre productos y estrategias corporativas.

Caso Real

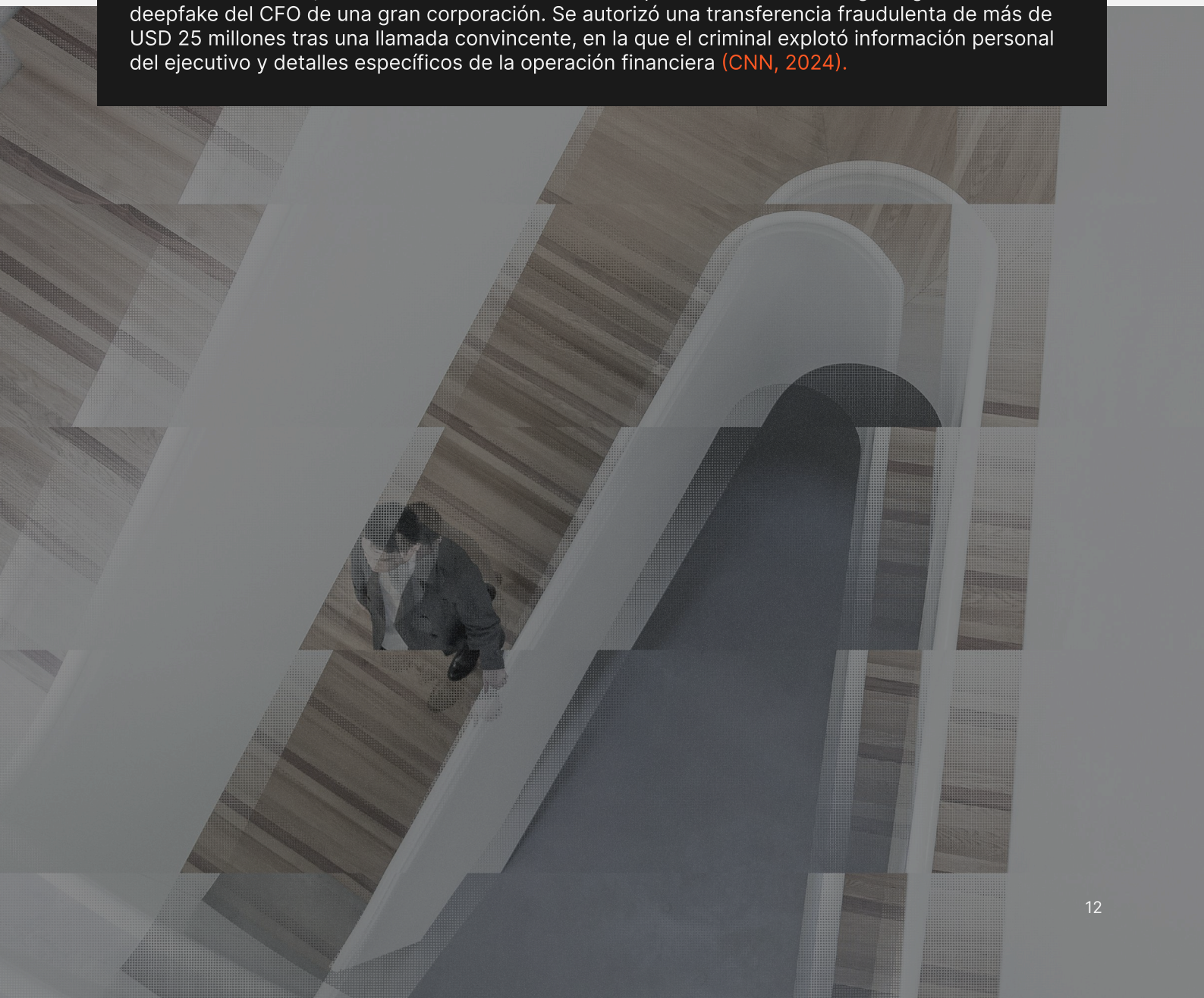
Un ejemplo reciente es el ataque contra la empresa Binance, donde criminales utilizaron un holograma deepfake del ejecutivo Patrick Hillman para manipular información sobre supuestos proyectos financieros inexistentes, afectando la credibilidad de la exchange (Bitdefender, 2023).

Escenario #2 Ingeniería social avanzada en el entorno corporativo

En esta situación, el criminal utiliza deepfakes para ejecutar ataques convincentes de ingeniería social dirigidos a altos ejecutivos o empleados estratégicos. El atacante primero realiza investigaciones detalladas sobre ejecutivos, recopilando datos personales en redes sociales y materiales audiovisuales para entrenar modelos avanzados. Después de obtener información sobre eventos corporativos recientes, como nuevas joint ventures o fusiones, el atacante crea deepfakes audiovisuales del CEO o del director financiero, utilizando audio clonado para realizar llamadas fraudulentas.

Ejemplo técnico y caso real

En febrero de 2024, criminales llevaron a cabo un ataque sofisticado en Hong Kong utilizando un deepfake del CFO de una gran corporación. Se autorizó una transferencia fraudulenta de más de USD 25 millones tras una llamada convincente, en la que el criminal explotó información personal del ejecutivo y detalles específicos de la operación financiera (CNN, 2024).



Escenario #3

Ataques de ingeniería social contra instituciones financieras

Aquí, el atacante utiliza deepfakes de audio para comprometer sistemas de autenticación por voz en bancos e instituciones financieras. Después de comprar datos personales robados en la dark web (nombre, número de identificación, números de cuentas bancarias) y recopilar muestras audiovisuales de las víctimas en redes sociales, el criminal crea clones de voz para engañar los sistemas de autenticación y obtener acceso a las cuentas financieras de los objetivos.

Ejemplo técnico y caso real

Aunque no existen registros públicos detallados de este escenario exacto, se han identificado técnicas similares en ataques contra instituciones financieras utilizando clonación de voz para engañar a empleados en centros de atención, permitiendo transferencias fraudulentas y acceso indebido a cuentas, demostrando la viabilidad técnica (Fortune, 2024; BleepingComputer, 2024).

Escenario #4

Responsabilidad corporativa y fraudes mediante deepfake

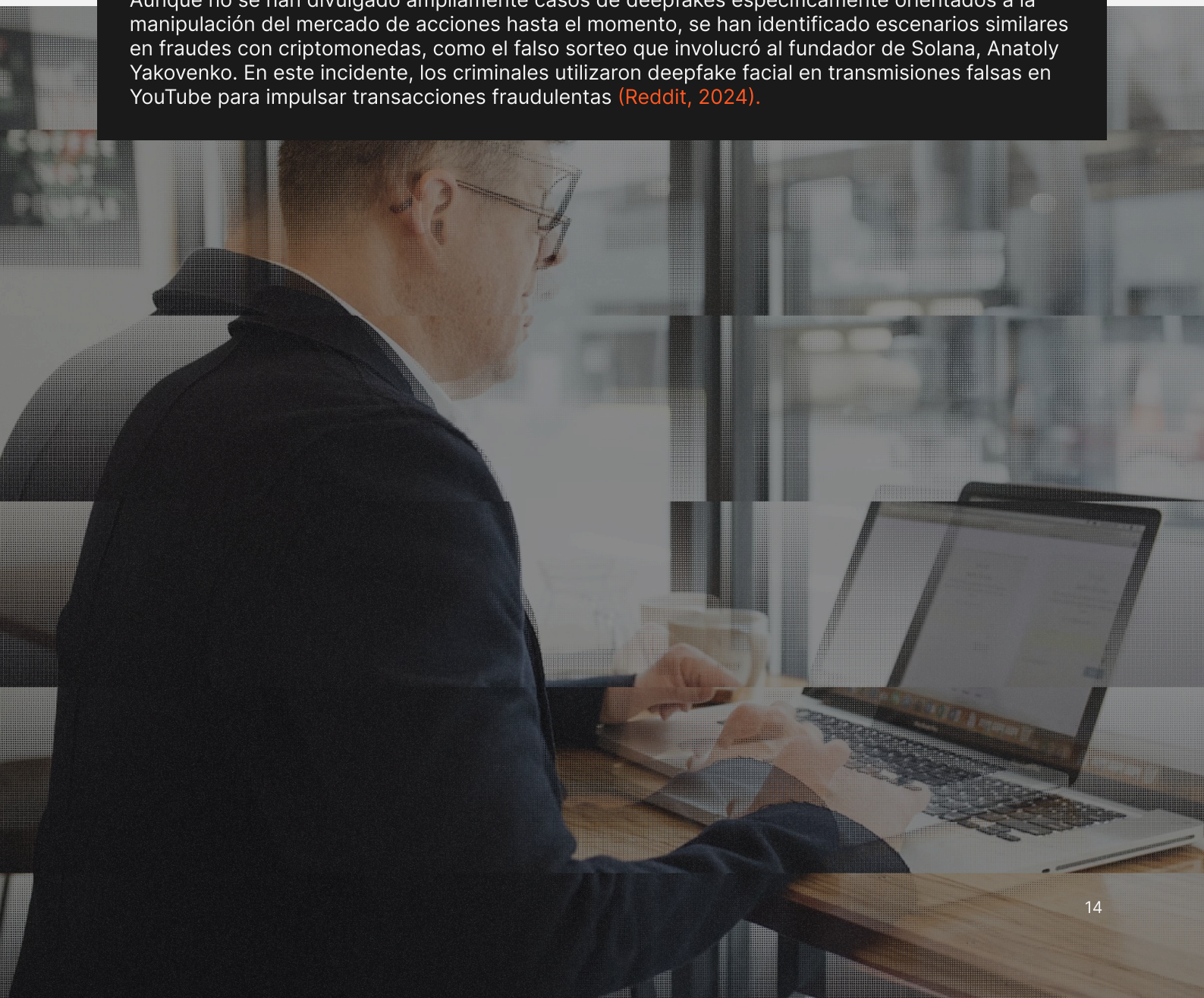
En este escenario, los criminales crean videos deepfakes realistas que sugieren fallos o accidentes graves relacionados con productos o servicios corporativos, con el objetivo de obtener compensaciones financieras fraudulentas. El criminal realiza investigaciones detalladas sobre incidentes reales anteriores y crea una narrativa visual falsa, utilizando modelos avanzados de intercambio facial para simular víctimas ficticias.

Escenario #5 Manipulación de mercado y acciones mediante deepfake

En este escenario, los atacantes utilizan deepfakes para manipular precios de acciones en el mercado financiero. El criminal adquiere previamente una posición accionaria y luego crea y distribuye contenido falso en foros populares como Reddit o Stockaholics. Deepfakes convincentes del CEO o de ejecutivos hablando sobre eventos financieros falsos inducen a los inversores a comprar o vender acciones rápidamente, provocando grandes variaciones en el precio del activo.

Ejemplo técnico y caso real

Aunque no se han divulgado ampliamente casos de deepfakes específicamente orientados a la manipulación del mercado de acciones hasta el momento, se han identificado escenarios similares en fraudes con criptomonedas, como el falso sorteo que involucró al fundador de Solana, Anatoly Yakovenko. En este incidente, los criminales utilizaron deepfake facial en transmisiones falsas en YouTube para impulsar transacciones fraudulentas (Reddit, 2024).



Cómo identificar y mitigar deepfakes: una guía técnica para empresas e individuos

¿Es posible educar al público para identificar deepfakes?

La educación y concienciación de la sociedad son cruciales para combatir eficazmente la propagación de medios sintéticos falsos. Sin embargo, el interés y el compromiso público con este aprendizaje pueden ser limitados.

Investigaciones realizadas por la PAI indican que las personas prefieren no ser instruidas de forma directa o condescendiente sobre qué es verdadero o falso; en cambio, prefieren aprender a través de contextos y herramientas que les permitan verificar la autenticidad de los medios de manera independiente.

Aun así, es posible capacitar a los individuos para identificar deepfakes enseñándoles a buscar señales técnicas claras, tales como:

Señales que indican que un video o imagen podría ser un deepfake:

→ Zonas borrosas o desenfocadas en el rostro, distintas del resto de la imagen.

→ Cambios repentinos en el tono de piel o en la iluminación alrededor de la cara.

→ Bordes duplicados (por ejemplo: doble mentón, cejas duplicadas).

→ Ausencia o exceso de parpadeos; movimientos faciales artificiales.

→ Problemas de sincronización labial y movimientos no naturales.

→ Inconsistencias entre el fondo y el objeto principal de la escena.

Señales que indican que un audio podría ser un deepfake:

→ Frases entrecortadas o fuera del ritmo natural.

→ Cambios bruscos en la entonación de la voz.

→ Uso inusual de palabras o frases que el supuesto interlocutor normalmente no diría.

→ Contexto fuera de una conversación habitual o incapacidad para responder preguntas simples relacionadas.

Señales que indican que un texto podría estar manipulado o ser fraudulento:

- → Errores ortográficos y gramaticales inusuales.
- → Falta de coherencia lógica y fluidez.
- → Correo electrónico o número desconocido o sospechoso.
- → Expresiones o vocabulario inusuales para el supuesto remitente.
- → Mensaje fuera de contexto o inesperado.

Sin embargo, a medida que la tecnología avanza, distinguir medios manipulados de los reales se convierte en una tarea progresivamente más compleja. Los deepfakes evolucionan rápidamente, impulsados por avances constantes en modelos generativos y técnicas sofisticadas de inteligencia artificial (GANs, autoencoders, entre otros).

Por lo tanto, depender exclusivamente de la percepción humana para detectar estos contenidos fraudulentos ya no es suficiente.

Para aumentar la eficacia de la identificación, es necesario integrar enfoques técnicos sólidos de análisis forense digital y algoritmos automatizados.

¿Cómo podemos mejorar la capacidad de detección?

Actualmente, herramientas como Reality Defender, Microsoft Video Authenticator, Deepware AI y frameworks académicos como FaceForensics++ están disponibles para analizar deepfakes con precisión cada vez mayor. Estos sistemas aplican análisis detallados utilizando algoritmos de machine learning entrenados específicamente para identificar características únicas, como inconsistencias en la iluminación, expresiones faciales no naturales o artefactos digitales dejados durante el proceso de manipulación.

Sin embargo, ninguna herramienta es totalmente infalible, especialmente frente a deepfakes altamente sofisticados y técnicamente avanzados. La eficacia de la detección depende directamente del entrenamiento continuo, los ajustes constantes de los modelos y la exposición permanente de las herramientas a nuevos ejemplos manipulados.

Además, las herramientas mencionadas no realizan un monitoreo avanzado para detectar fraudes que utilicen la imagen de ejecutivos en redes sociales, sitios web y plataformas de streaming, algo esencial para obtener visibilidad sobre las amenazas.

¿Cómo se conectan estas amenazas con la protección de marca y el fraude?

Las campañas con deepfakes que involucran a ejecutivos y figuras públicas no afectan únicamente a las víctimas directas. Cuando ocurren, los efectos se extienden a toda la estructura de la empresa: impactan la reputación institucional, afectan las relaciones con stakeholders y, muchas veces, se traducen en pérdidas financieras. El uso indebido de la imagen de líderes organizacionales no es solo una violación de identidad individual, sino también una puerta de entrada para fraudes de alcance corporativo.

Uso de identidad de ejecutivos en promociones falsas o campañas engañosas

En este tipo de fraude, un ejecutivo conocido aparece en video o audio promoviendo inversiones, anunciando asociaciones o ofreciendo ventajas que, en realidad, nunca existieron. En muchos casos, este contenido es patrocinado y difundido en redes sociales como si formara parte de la comunicación oficial de la empresa.
La consecuencia directa es el deterioro de la marca. Colaboradores, clientes o socios pueden asociar el contenido falso con la comunicación legítima de la organización, generando confusión y pérdida de confianza. Incluso después de contener el incidente, el esfuerzo por reconstruir la credibilidad puede ser largo y costoso.

Es importante destacar que los delincuentes no necesitan necesariamente utilizar el rostro de la víctima; características físicas distintivas o elementos marcarios del estilo personal también pueden inducir a una asociación errónea.

Vinculación entre ataques dirigidos y incidentes de marca

Otro punto de atención es la superposición entre ataques de ingeniería social y abusos de marca. Los criminales han combinado deepfakes de ejecutivos con dominios falsos, correos electrónicos visualmente legítimos y perfiles clonados en redes sociales para crear contextos creíbles. El uso de identidad corporativa en estos ataques —logos, nombres, slogans e incluso elementos del tono de voz institucional— refuerza la ilusión de autenticidad.

Para los equipos de seguridad de la información, esto representa un desafío que va más allá de la detección de amenazas técnicas. Implica monitorear señales de abuso de imagen y narrativa corporativa en múltiples entornos, y responder rápidamente para evitar que una acción aislada escale hacia una crisis institucional.

La respuesta a este tipo de ataque debe ir más allá del monitoreo: exige una intervención directa en los entornos donde se están difundiendo los contenidos fraudulentos. Es en este contexto donde entra el concepto de takedown: un proceso de eliminación de contenidos o activos maliciosos, solicitado ante plataformas, proveedores de servicios, redes sociales, registradores de dominios, entre otros intermediarios.

¿Qué puede ser eliminado mediante el takedown?

En el caso de fraudes con deepfakes, el takedown se convierte en una herramienta esencial cuando el contenido manipulado infringe las políticas de uso de la plataforma donde está alojado. Perfiles falsos de ejecutivos en redes sociales, videos alojados en sitios de terceros, anuncios engañosos, páginas de phishing con identidad visual de la empresa —todos estos elementos son, en general, susceptibles de ser eliminados.

La ley federal estadounidense Take It Down Act exige que las plataformas dedicadas principalmente a la distribución de contenido generado por usuarios cuenten con un canal específico y ágil para denunciar y eliminar, en un plazo máximo de 48 horas, representaciones visuales íntimas, ya sean auténticas o generadas por IA, como los deepfakes. La ley no exige que la imagen sea pornográfica ni que muestre el rostro de la persona; cualquier rasgo que permita identificar al individuo —como una marca de nacimiento— puede ser suficiente para justificar la solicitud de eliminación conforme a esta legislación.

Por otro lado, **no todo tipo de contenido generado con IA puede ser eliminado con la misma facilidad**. Contenidos que simulan contextos ficticios sin infringir directamente una marca registrada, por ejemplo, pueden escapar de los criterios tradicionales de moderación. En estos casos, el rol de la seguridad de la información es evaluar el riesgo reputacional, buscar evidencias técnicas de manipulación y activar canales alternativos —como comunicación institucional, área jurídica y gestión de crisis.

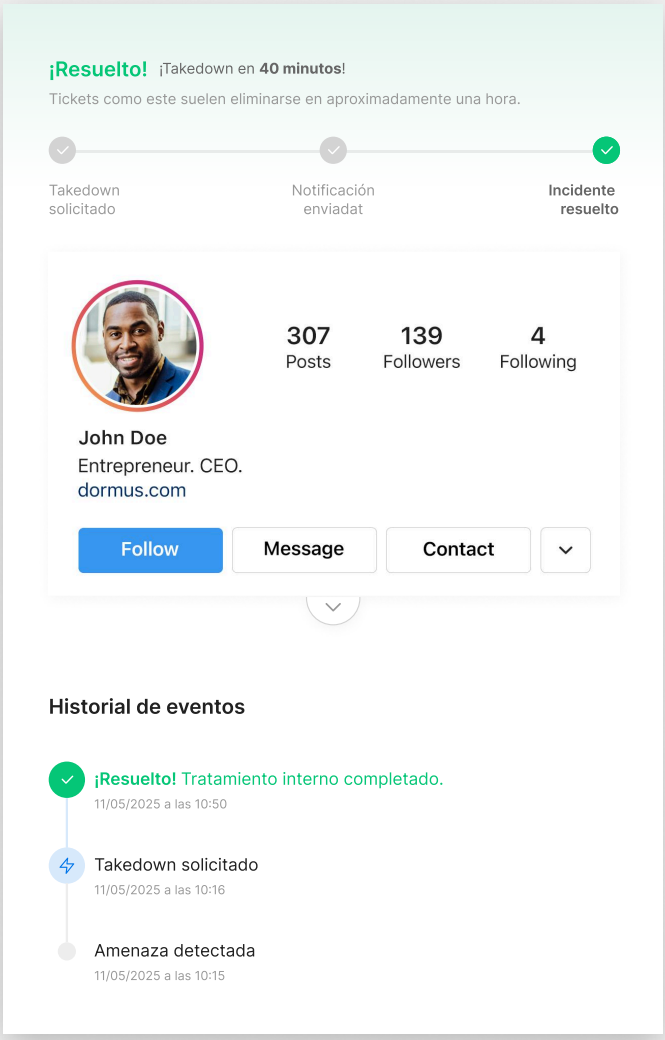
Qué puede (o no) ser eliminado en un fraude con deepfake

Elemento utilizado en el fraude	Tipo de afectación	Tipo de afectación
Correo electrónico a nombre del ejecutivo con dominio similar al oficial	Corporativo	✓ **Sí** El dominio puede ser eliminado
Perfil falso con nombre y foto del ejecutivo en LinkedIn o Instagram	Personal (impacto corporativo)	✓ **Sí** El perfil puede ser eliminado
Página falsa con identidad visual de la empresa (logos, colores, nombre comercial)	Corporativo	✓ **Sí** Puede ser eliminada por violación de marca
Video deepfake del ejecutivo promoviendo una inversión falsa con uso de la marca	Ambos	✓ **Sí** Pasible de eliminación (plataformas y anuncios)
Video deepfake del ejecutivo sin uso de marca, pero con apariencia realista	Personal	✓ **Sí** Pasible de eliminación
Imagen o video con escena íntima cuya difusión no fue autorizada, sea auténtica o generada	Ambos	✓ **Sí** Conforme al Take It Down Act, debe ser eliminado en un plazo de 48 horas a partir de 2026
Video en plataforma extranjera con simulación genérica de ejecutivo, sin marca	Ambos	× **No** Difícil de eliminar sin violación explícita
Dominio falso simulando subdominio oficial (ej: financiero.suempresa.online)	Corporativo	✓ **Sí** Puede ser eliminado por abuso de marca
Audio clonado del ejecutivo enviado por app de mensajería	Personal (impacto interno)	⚠ **No directamente** Difícil de rastrear o eliminar
Video falso publicado en grupo cerrado (WhatsApp, Telegram, etc.)	Ambos	⚠ **No directamente** Pero los canales pueden ser monitoreados e investigados
Publicación falsa simulando declaración institucional firmada por un ejecutivo	Corporativo	✓ **Sí** Puede ser eliminada por violación de imagen y marca

La plataforma Axur viabiliza el proceso de identificación y solicitud de takedown a escala, con soporte para diferentes tipos de activos y canales.

En escenarios donde la velocidad de respuesta es determinante —como una campaña de fraude utilizando la imagen del CEO en un perfil falso— el tiempo de reacción puede ser el factor decisivo entre contener un incidente o permitir que escale.

Para los equipos de seguridad, esta capacidad representa no solo una reacción más rápida, sino también un punto de integración con otras áreas de la empresa, ampliando la visibilidad y la coordinación en incidentes de impacto externo.



Conclusión

Como vimos a lo largo de este material, la superficie de ataque no se limita a los sistemas. Se extiende a la identidad institucional y a la presencia pública de los líderes de la organización, que, cuando son manipuladas, se convierten en herramientas para fraudes con potencial de causar pérdidas financieras, crisis de imagen e impactos estratégicos.

La tendencia es que estos ataques se vuelvan aún más convincentes. Tecnologías de clonación de voz en tiempo real, avatares interactivos y la integración entre deepfakes y modelos de lenguaje (LLMs) ya están comenzando a hacer posible la creación de ejecutivos sintéticos capaces de dialogar con las víctimas de manera personalizada y adaptable. La escalabilidad de estos ataques exigirá, cada vez más, una respuesta coordinada entre los equipos de seguridad, jurídico, comunicación y gobernanza.

La plataforma Axur fue desarrollada para apoyar a las empresas en este nuevo escenario. Automatizamos la detección de amenazas externas, el monitoreo de perfiles falsos y la ejecución de takedown a escala, ofreciendo visibilidad y agilidad en la contención de fraudes que involucran la imagen de ejecutivos y la identidad de la marca.

Si su empresa ya está enfrentando incidentes de este tipo —o desea prepararse para enfrentarlos con confianza— estamos a su disposición para apoyar su estrategia de defensa.