ANÁLISIS SOBRE LA CIBERDELINCUENCIA

BOLETÍN

EN ESTA EDICIÓN

DEEPFAKES Y FRAUDE DIGITAL: LA NUEVA ERA DE LAS ESTAFAS EN LÍNEA

DIRECCIÓN DE CIBERDELITOS
MINISTERIO DEL INTERIOR

Boletín de Análisis sobre la Ciberdelincuencia: Deepfakes y fraude digital: la nueva era de las estafas en línea

Ministerio del Interior
Subsecretaría de Combate al Delito
Dirección de Ciberdelitos

Presidente de la República
MAGISTER DANIEL ROY-GILCHRIST NOBOA AZÍN

Ministro del Interior
Sr. JOHN REIMBERG OVIEDO

Subsecretario de Combate al Delito
TENIENTE CORONEL (SP) LUIS FERNANDO PÉREZ DÁVILA

Director de Ciberdelitos, Encargado
MAGÍSTER JORGE NÉJER GUERRERO

Responsables y Colaboradores

Redacción técnica del documento:
MAGÍSTER GABRIEL REINOSO MARTÍNEZ, Analista de Ciberdelitos
MAGÍSTER CARLOS SIMBAÑA COBA, Analista de Ciberdelitos
INGENIERO CÉSAR TRELLES SEGOVIA, Analista de Ciberdelitos
INGENIERO FABIÁN SILVA TOLEDO, Analista de Ciberdelitos

Revisión técnica del documento:
INGENIERO FREDDY GALLARDO SOSA, Especialista de Ciberdelitos

Redacción y compilación:
MAGÍSTER DUVAL MONTATIXE CAIZALUISA, Analista de Ciberdelitos

Edición y Adaptación:
MAGÍSTER DUVAL MONTATIXE CAIZALUISA, Analista de Ciberdelitos

MARZO DE 2026

Tabla de contenido

Contenido

Resumen ejecutivo

El Boletín de Análisis sobre la Ciberdelincuencia **“Deepfakes y fraude digital: la nueva era de las estafas en línea**”, analiza el impacto de los deepfakes como modalidad de fraude digital en el Ecuador. Estos contenidos manipulados, generados mediante inteligencia artificial, permiten la creación de audios, videos e imágenes falsas con alto realismo, lo que facilita la suplantación de identidad y la manipulación emocional de las víctimas.

El documento analiza cómo la evolución reciente de la inteligencia artificial ha transformado las dinámicas delictivas tradicionales; esta evolución ha debilitado los mecanismos convencionales de verificación de identidad y ha fortalecido esquemas de ingeniería social, orientados a manipular la confianza, las emociones y la toma de decisiones de las víctimas, en función de sus factores de vulnerabilidad. Asimismo, identifica que las estafas con deepfakes responden a una secuencia estructurada y planificada, caracterizada por la recolección previa de información pública, la generación de contenido falso y la obtención de beneficios económicos o el acceso a información sensible.

El análisis del contexto nacional, a partir de la comparativa de noticias del delito, evidencia que entre los años 2023 y 2025, se registró una reducción de 6.474 a 5.573 casos delictivos, lo que supone una disminución moderada de los delitos relacionados con deepfakes. No obstante, este comportamiento no implica la contención del fenómeno, sino su evolución hacia modalidades más sofisticadas, potenciadas por el uso de inteligencia artificial y esquemas híbridos de engaño, en los que los deepfakes adquieren un rol cada vez más relevante.

El boletín destaca, además, los impactos económicos, sociales, políticos y reputacionales asociados a este tipo de conductas ilícitas, que utilizan las tecnologías de la información y comunicación como medio y fin para el cometimiento de delitos, incluyendo afectaciones al sector financiero, instituciones públicas, personas naturales y a la confianza en las comunicaciones digitales legítimas.

Finalmente, el documento propone un conjunto de medidas de prevención y mitigación, orientadas tanto a la ciudadanía como a las instituciones, que incluyen la alfabetización digital, el fortalecimiento de protocolos de verificación de identidad, la coordinación interinstitucional y el uso de herramientas especializadas para la detección de deepfakes.

Introducción

En el entorno digital actual, los mecanismos tradicionales de validación basados en la voz, el texto y la imagen han perdido confiabilidad como medios de verificación de identidad. Este cambio responde al desarrollo de técnicas capaces de generar datos sintéticos, conocidas como deepfakes, que elevan la sofisticación de las estafas digitales y configuran una modalidad cada vez más utilizada en ciberdelincuencia basada en la suplantación de identidad.

Estas técnicas permiten la creación de audios y videos falsos con un alto nivel de realismo, capaces de reproducir la voz, la imagen y los patrones comunicativos de personas reales, lo que dificulta la verificación de la autenticidad del contenido. (Citron & Chesney , 2019) (Kietzmann, McCarthy, Lee, & Kietzmann, 2020).

Desde el análisis del comportamiento del delito, las estafas con deepfakes superan la condición de hechos aislados y responden a un esquema organizado, caracterizado por planificación, uso estratégico de recursos y adaptación tecnológica. Tal como señala la doctrina especializada en ciberdelincuencia, este tipo de delitos responde a esquemas secuenciales y repetibles, orientados a maximizar la efectividad del engaño y reducir las posibilidades de detección temprana.

El presente documento analiza la forma en que se ejecutan las estafas con deepfakes e identifica los factores que facilitan estos delitos, las víctimas y la normativa vigente en el contexto nacional.

“El fraude digital trasciende hechos aislados y responde a patrones definidos.”

Objetivo

Analizar el uso de deepfakes como una modalidad de fraude digital en Ecuador, considerando las estrategias de manipulación empleadas, los factores de vulnerabilidad de las víctimas y los patrones de comportamiento de los estafadores, con el fin de orientar acciones de prevención y mitigación frente a este tipo de ciberdelincuencia.

Deepfakes y su uso en estafas digitales

En el actual entorno digital los deepfakes se definen como contenidos manipulados mediante inteligencia artificial, capaces de generar videos, audios e imágenes falsos con apariencia real.

Esta modalidad se ha convertido en un mecanismo clave para la suplantación de identidad, representando un desafío creciente para el combate a la ciberdelincuencia. En la práctica, permite clonar rostros y voces a partir de contenidos públicos para engañar a las víctimas, y se utiliza en estafas y extorsiones mediante la simulación de comunicaciones de personas o instituciones, con el fin de inducir la entrega de dinero, información sensible o la ejecución de acciones bajo presión, generando impactos económicos y emocionales.

Inicialmente asociados a usos indebidos en contenidos difundidos sin consentimiento, los deepfakes han evolucionado hacia instrumentos de desinformación digital que facilitan la difusión de contenidos falsos con apariencia de veracidad en internet y redes sociales.

Los deepfakes representan una amenaza relevante para la sociedad, por sus implicaciones sociales, políticas y económicas, al ser utilizados para difundir información falsa, afectar la reputación, cometer fraudes, influir en la toma de decisiones, manipular la opinión pública, alterar mercados financieros o afectar las relaciones internacionales.

Los deepfakes pueden presentarse en distintos formatos y ser utilizados en esquemas de fraude y manipulación. En el caso de los audios, permiten replicar la voz de una persona para simular llamadas o mensajes falsos, con el fin de simular la identidad de familiares, autoridades o representantes institucionales e inducir decisiones bajo engaño. En el ámbito visual, los deepfakes de video permiten la simulación de videollamadas o contenidos donde se aparenta la participación de una persona real, reforzando la credibilidad del engaño y facilitando la manipulación de las víctimas (LISA INSTITUTE, 2026).

De forma complementaria, los delincuentes utilizan perfiles falsos en redes sociales para construir identidades aparentes, generar confianza y ejecutar estafas, especialmente en entornos de comercio digital mediante solicitudes de pagos anticipados. Estas modalidades

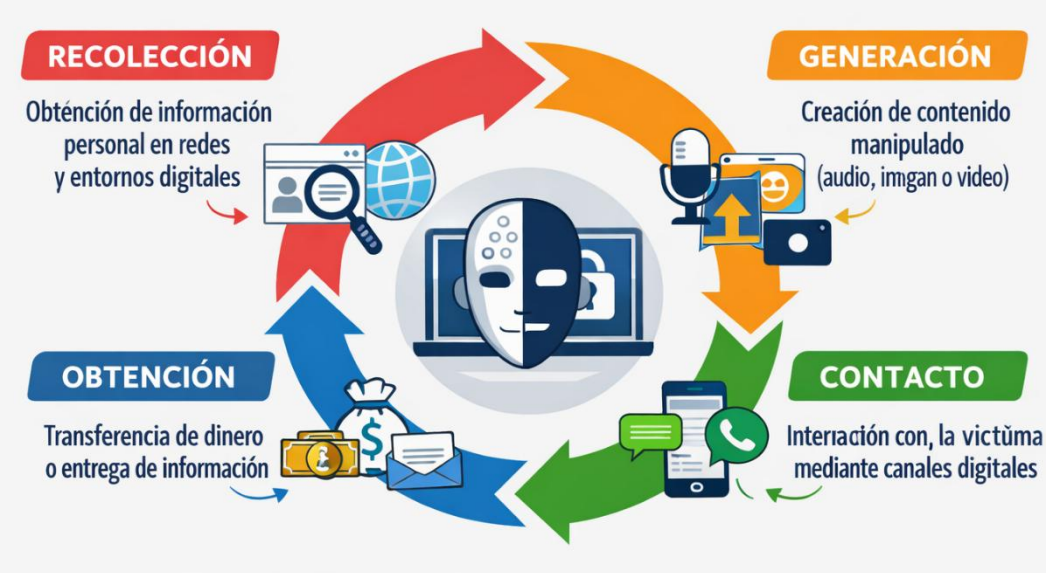
pueden combinarse en esquemas más complejos, integrando audio, video y contenido automatizado para incrementar la efectividad del engaño y maximizar el impacto sobre las víctimas (INCIBE, 2023).

> **“El realismo del contenido manipulado aumenta la posibilidad de engaño.”**

Mecanismo de operación de las estafas con deepfakes

El proceso de las estafas con deepfakes se desarrolla a través de fases definidas que estructuran el engaño de manera progresiva, desde la obtención de información, la generación de contenido manipulado y el contacto con la víctima, hasta la obtención del beneficio ilícito, como se presenta en el siguiente esquema:

Figura 1
Proceso de estafas con deepfakes



Elaboración: *Dirección de Ciberdelitos – Ministerio del Interior*

A partir de este esquema, se presenta a continuación el desarrollo de cada una de las fases que conforman el proceso, con el fin de comprender su funcionamiento y las dinámicas utilizadas.

Recolección de información pública de la víctima

La primera fase del modus operandi consiste en la obtención sistemática de información sobre la víctima, principalmente a partir de fuentes abiertas: redes sociales, plataformas de mensajería, repositorios de video y entornos profesionales digitales constituyen espacios habituales de extracción de datos personales y de material audiovisual susceptible de ser reutilizado.

El entorno digital facilita este tipo de delitos, ya que la información disponible en internet amplía las oportunidades de los delincuentes. En este contexto, la selección de la víctima se basa en su nivel de exposición y en la disponibilidad de contenido que pueda ser utilizado para el engaño, lo que incrementa la probabilidad de éxito.

Generación de contenido falso mediante inteligencia artificial

Una vez recopilada la información necesaria, el ciberdelincuente procede a la generación de contenido falso mediante sistemas de inteligencia artificial, con el objetivo de suplantar de manera creíble la identidad de la víctima. A partir de muestras reales de voz e imagen, los modelos generativos permiten producir audios y videos sintéticos que replican rasgos físicos, entonación, gestos y patrones comunicativos con un alto grado de realismo. Esta técnica permite suplantar la identidad de una persona de forma más convincente y dificulta verificar si un audio o video es real, así como identificar a los responsables (Europol, 2022).

Contacto directo y manipulación de la víctima

La fase de contacto directo representa la materialización del engaño. En esta etapa, el ciberdelincuente establece comunicación con la víctima o con personas de su entorno a través de canales digitales de uso cotidiano, como llamadas telefónicas, aplicaciones de mensajería instantánea o videollamadas. Este mecanismo se apoya en técnicas clásicas de ingeniería social, tales como la construcción de escenarios de urgencia, autoridad o necesidad emocional, lo que incrementa la apariencia real del mensaje.

Diversos estudios han demostrado que la exposición a contenido sintético altamente realista disminuye la eficacia de los mecanismos racionales de verificación, incluso en usuarios con experiencia digital; lo cual evidencia que la vulnerabilidad ante este tipo de estafas se explica por el uso habitual de los canales digitales, independientemente del criterio individual de la víctima (Vaccari & Chadwick, 2020).

Obtención de beneficios económicos o información sensible

La fase final tiene como propósito la obtención de un beneficio ilícito, ya sea económico o informacional. La víctima, inducida al error, realiza actos de disposición patrimonial, como transferencias de dinero, pagos o facilita credenciales de acceso e información sensible

susceptible de reutilización delictiva. El daño ocasionado trasciende el perjuicio económico inmediato, extendiéndose a afectaciones emocionales, pérdida de confianza y sensación de inseguridad digital.

Estrategias de manipulación emocional y psicológica

La manipulación emocional es el eje central del fraude digital porque permite a los ciberdelincuentes influir en la toma de decisiones de las víctimas, debilitando su capacidad de análisis crítico. Estas estrategias se potencian mediante el uso de técnicas de engaño, deepfakes y la personalización del ataque a partir de información obtenida en entornos digitales.

Los desencadenantes emocionales más utilizados por los ciberdelincuentes son:

- **Explotación de la confianza:** Los atacantes se hacen pasar por personas cercanas o de confianza para generar seguridad y evitar que la víctima verifique la información.
- **Inducción del miedo y la urgencia**: Generan escenarios de crisis y pánico (emergencias familiares, amenazas legales, bloqueos de cuentas, entre otros) para impulsar decisiones apresuradas.
- **Manipulación de la empatía y la solidaridad:** Apelan a la compasión, afecto o responsabilidad moral, simulando situaciones de necesidad extrema; esta estrategia es común en estafas dirigidas a entornos familiares.
- **Uso de autoridad y jerarquía:** Suplantan a figuras de poder como jefes, directivos, funcionarios públicos, para imponer órdenes o solicitudes económicas, lo que reduce la tendencia a cuestionar la legitimidad del mensaje.
- **Sobrecarga cognitiva:** La víctima recibe múltiples estímulos simultáneos como mensajes, llamadas, audios, documentos, entre otros, para generar confusión y reducir la capacidad de análisis.
- **Curiosidad:** Los ciberdelincuentes emplean mensajes llamativos que incitan a acceder a enlaces maliciosos.
- **Recompensas:** Prometen beneficios falsos para obtener datos financieros de las víctimas (WatchGuard, 2025).

Factores de vulnerabilidad de las víctimas

Los fraudes digitales potenciados por deepfakes explotan factores que aumentan el riesgo en las personas, especialmente en entornos digitales:

- **Exposición excesiva de información personal:** publicación de datos, rutinas o relaciones en redes sociales que facilita la personalización del engaño.

- **Confianza elevada en entornos digitales:** tendencia a asumir que los mensajes o perfiles son legítimos, incluso ante solicitudes inusuales.
- **Vulnerabilidad emocional y estrés:** situaciones de presión, miedo o preocupación que reducen la capacidad de análisis y favorecen decisiones impulsivas.
- **Bajo nivel de conocimiento digital:** dificultad para identificar señales de alerta, especialmente frente a tecnologías como deepfakes.
- **Influencia de figuras de autoridad:** mayor probabilidad de obedecer solicitudes cuando aparentan provenir de jefes, autoridades o instituciones.
- **Dependencia tecnológica:** necesidad de responder de forma inmediata que limita la verificación de la información.
- **Aislamiento digital de la víctima:** interacción en canales privados que impide contrastar la información con otras personas.

Patrones de comportamiento de los estafadores

El uso de deepfakes en estafas digitales responde a patrones conductuales, estructurados y repetibles, entre los cuales se identifican los siguientes:

- **Recolección previa de información:** obtienen datos de la víctima en redes sociales u otras fuentes para personalizar el engaño.
- **Creación de identidades creíbles:** construyen o clonan perfiles que aparentan ser personas reales o de confianza.
- **Uso de emociones desde el primer contacto:** generan miedo, urgencia o confianza para reducir la capacidad de análisis.
- **Desarrollo progresivo del engaño:** avanzan por etapas, desde un contacto inicial hasta una solicitud final de dinero o información.
- **Uso de canales privados de comunicación:** prefieren medios como mensajes directos o llamadas para evitar verificación con terceros.
- **Presión de tiempo:** exigen respuestas inmediatas para impedir que la víctima piense o confirme la información.
- **Adaptación del discurso:** ajustan el mensaje según la reacción de la víctima para mantener la credibilidad.
- **Corte de comunicación tras lograr el objetivo:** desaparecen o bloquean contacto una vez obtenido el beneficio.

Relevancia en el contexto nacional

Para el año 2025, el Ecuador registra una población total estimada de 18.103.660 habitantes; el 80,1 % de la población utiliza Internet, y el 59,3 % posee un teléfono inteligente (INEC, 2025). Asimismo, Guayas y Pichincha concentran el 58 % de las conexiones a internet del país, según el informe Ecuador Digital, publicado en noviembre de 2025 por la consultora Mentinno, especializada en analítica, negocios y talento. Este escenario evidencia una alta disponibilidad de conectividad en la población, que ofrece oportunidades para el aprendizaje, la comunicación y el entretenimiento; a la vez, amplía la exposición a riesgos en el uso cotidiano de las tecnologías de la información y la comunicación.

El uso creciente de las redes sociales también resulta significativo. Según el mismo informe, las conexiones a Google y YouTube superan los 18,5 millones, mientras que TikTok registra 17,3 millones de cuentas, lo que las posiciona entre las plataformas con mayor presencia en el país, reflejando un incremento del 20,8 % respecto del año 2024. Este panorama evidencia la integración de plataformas digitales en actividades cotidianas, laborales y comunicacionales (Mentinno, 2025).

En este contexto, la digitalización de las interacciones personales, laborales e institucionales ha convertido a la voz, la imagen y los mensajes digitales en elementos habituales de confianza.

Esta situación adquiere especial relevancia en el país debido al uso extendido de herramientas digitales para trámites, comunicaciones laborales y transacciones económicas, así como a brechas en alfabetización digital que incrementan la exposición a fraudes y engaños.

Desde una perspectiva institucional, estas modalidades de fraude se caracterizan por su rápida adaptación, expansión y por las dificultades que generan para la identificación de responsables y la verificación de evidencias digitales.

En el marco normativo nacional, el Código Orgánico Integral Penal, en su artículo 212, tipifica la suplantación de identidad en los siguientes términos: “La persona que de cualquier

forma suplante la identidad de otra para obtener un beneficio para sí o para un tercero, en perjuicio de una persona, será sancionada con pena privativa de libertad de uno a tres años". En este sentido, el análisis de la evolución de las noticias del delito registradas por la Fiscalía General del Estado en materia de suplantación de identidad durante el período del año 2023 al año 2025 evidencia una disminución sostenida del 13,9 % respecto del año 2023, conforme se observa en la siguiente tabla:

Tabla 1. *Noticias del delito Fiscalía General del Estado, Suplantación de identidad*

Año	Noticias del delito Suplantación de identidad	Variación porcentual
2023	6.474	0 %
2024	5.920	-8,6 %
2025	5.573	-5,9 %

Elaboración*: Dirección de Ciberdelitos - Ministerio del Interior.* ***Fuente:*** *Fiscalía General del Estado*

Si bien las noticias del delito por suplantación de identidad presentan una disminución moderada, este comportamiento puede interpretarse como una reconfiguración de sus formas de ejecución hacia modalidades más sofisticadas, mediante el uso de herramientas tecnológicas y esquemas de engaño más complejos, que combinan mecanismos tradicionales con herramientas basadas en inteligencia artificial.

Por su parte, el Código Orgánico Integral Penal, en el artículo 186, numerales 1 y 2, tipifica determinadas modalidades de estafa vinculadas al uso fraudulento de medios de pago y dispositivos electrónicos. En este marco, el análisis de la evolución de las noticias del delito registradas por la Fiscalía General del Estado respecto de estafas durante el período del año 2023 al año 2025 evidencia una tendencia creciente y sostenida, equivalente a un incremento aproximado del 43,1 % respecto al año 2023, como se muestra a continuación:

Tabla 2. *Noticias del delito Fiscalía General del Estado, Estafas numeral 1 y 2*

Año	Noticias del delito Estafas numeral 1 y 2	Variación porcentual
2023	965	0 %
2024	1.125	16,8 %
2025	1.381	22,8 %

Elaboración*: Dirección de Ciberdelitos – Ministerio del Interior.* ***Fuente:*** *Fiscalía General del Estado*

El crecimiento de las estafas evidencia que el fenómeno persiste y se intensifica. Este comportamiento refleja una mayor sofisticación en las formas de fraude utilizadas en el país.

En el ámbito económico e institucional, estas modalidades impactan a empresas, instituciones públicas y al sector financiero, generando pérdidas económicas y afectaciones reputacionales. Entre las principales amenazas se identifican la suplantación de identidades para validar instrucciones irregulares y la difusión de información falsa que puede afectar la estabilidad institucional y la toma de decisiones.

Además, estas prácticas generan incertidumbre sobre la autenticidad de audios, videos y mensajes, dificultando la verificación de la información que circula en canales digitales.

“La expansión digital también amplía la superficie de exposición al fraude.”

Los contenidos manipulados mediante deepfakes incorporan alteraciones de imagen, voz y video con un alto nivel de similitud con la realidad, afectando derechos como la imagen, la intimidad y la protección de datos personales. Estos contenidos pueden ser utilizados para extorsión, fraude, desinformación política, violencia de género digital y afectación a la reputación de personas y entidades públicas o privadas (Romero N., Sinaluisa S., & Freire , 2024).

En el contexto ecuatoriano, la existencia de dificultades para la persecución de delitos específicos vinculados a los deepfakes, pero también subrayan que estos pueden encuadrarse en figuras ya previstas en el Código Orgánico Integral Penal (COIP), como delitos contra la intimidad, violencia digital, fraude o delitos informáticos, dependiendo del caso concreto.

Impacto del Fenómeno en el contexto Nacional

En el contexto ecuatoriano, estas modalidades de fraude han evolucionado hacia esquemas más complejos que combinan tecnología y manipulación de información, generando nuevos desafíos para la detección, respuesta y verificación de contenidos.

Dimensión político-estratégica

En el ámbito electoral, se han identificado casos de manipulación de información mediante contenidos falsos que simulan declaraciones o acciones de actores políticos, lo que puede influir en la opinión pública y distorsionar la toma de decisiones ciudadanas. Asimismo, la generación de material audiovisual falso permite construir narrativas que afectan la percepción sobre la gestión pública y debilitan la credibilidad institucional.

Dimensión social

Uno de los impactos más sensibles se relaciona con el uso de estas tecnologías en contextos de violencia digital, especialmente en la generación de contenido no consentido o en la afectación de la reputación de personas, incluyendo mujeres y menores de edad. Estas prácticas amplifican daños emocionales y sociales, y evidencian la necesidad de fortalecer mecanismos de protección en entornos digitales.

Dimensión económica e institucional

En el ámbito económico, estas modalidades de fraude afectan a empresas, instituciones públicas y al sector financiero, mediante la suplantación de identidad para validar instrucciones irregulares, acceder a información sensible o ejecutar transacciones fraudulentas.

El sector financiero ha identificado un crecimiento sostenido de ataques que incorporan componentes de inteligencia artificial. Se estima que:

- Alrededor del 20 % de los ciberataques en el sector financiero ya incorporan elementos de IA (Asobanca, 2025).
- Las pérdidas por fraudes en transacciones digitales alcanzan aproximadamente USD 0,83 millones mensuales, con tendencia creciente (Maldonado, Sánchez, Ramirez, & Hallo, 2024).
- A nivel país, las pérdidas económicas asociadas al fraude digital pueden alcanzar hasta USD 600 millones anuales, evidenciando la magnitud del impacto en la economía nacional (Banco Mundial, 2025).

Adicionalmente, estas prácticas generan costos indirectos asociados a la gestión de incidentes, recuperación de sistemas y mitigación del daño reputacional, lo que obliga a las instituciones a incrementar sus inversiones en ciberseguridad y gestión de crisis.

Tabla 3. *Pérdidas económicas asociadas al fraude digital en el Ecuador*

Categoría de Pérdida	Monto / Indicador	Fuente Principal
Pérdida País (Anual)	$600 Millones (Máx.)	Banco Mundial
Costo Violación de Datos	$4.88 Millones (Promedio, 2025)	Costo de una filtración de datos Informe 2025 (IBM, 2025)
Aumento de Fraude IA	+115 % en el último año	(Kaspersky, 2025)
Fraudes Digitales (Mes)	$833.333 USD	El delito de fraude financiero en el Ecuador (Maldonado, Sánchez, Ramirez, & Hallo, 2024)

Elaboración*: Dirección de Ciberdelitos – Ministerio del Interior*

Impacto en la verificación de la información

Un efecto relevante es la creciente dificultad para distinguir entre contenidos reales y manipulados, lo que afecta procesos de validación en ámbitos judiciales, administrativos y comunicacionales. Esta situación incrementa los tiempos de respuesta institucional y complica la gestión de información en contextos de crisis. (Asobanca, 2025).

Erosión de la confianza en las comunicaciones digitales legítimas en Ecuador

En el contexto local, el mayor impacto de los deepfakes abarca tanto la desinformación como el fenómeno conocido como "Dividendo del mentiroso". Este concepto se refiere a situaciones en las que figuras públicas y actores políticos descalifican contenidos verdaderos, alegando que se trata de falsificaciones digitales.

En materia de rendición de cuentas, según investigaciones de la Universidad Politécnica Salesiana la presencia de medios sintéticos permite que actores estatales y privados siembren duda razonable sobre evidencia legítima (audios de WhatsApp, videos de vigilancia o grabaciones de llamadas), paralizando procesos judiciales y administrativos por la supuesta "dificultad probatoria" (Universidad Politécnica Salesiana, 2025).

La confianza en los medios de comunicación tradicionales, históricamente los "árbitros" de la verdad en Ecuador, ha sido severamente vulnerada. Los atacantes no solo crean noticias falsas, sino que alteran la percepción de legitimidad mediante el uso de identidades corporativas reconocidas.

La suplantación de identidad visual constituye otra modalidad de la que ha registrado varios casos en el Ecuador, con el uso de plantillas gráficas de medios de comunicación televisivos y la clonación de voces de presentadores icónicos para emitir comunicados de pánico financiero o decretos de excepción falsos. Esto genera un estado de "alerta constante" donde el ciudadano, ante la duda, reduce su capacidad de reacción incluso ante emergencias reales (Miranda Romero, 2025).

Caso documentado en Ecuador: suplantación de presentador de medio televisivo nacional

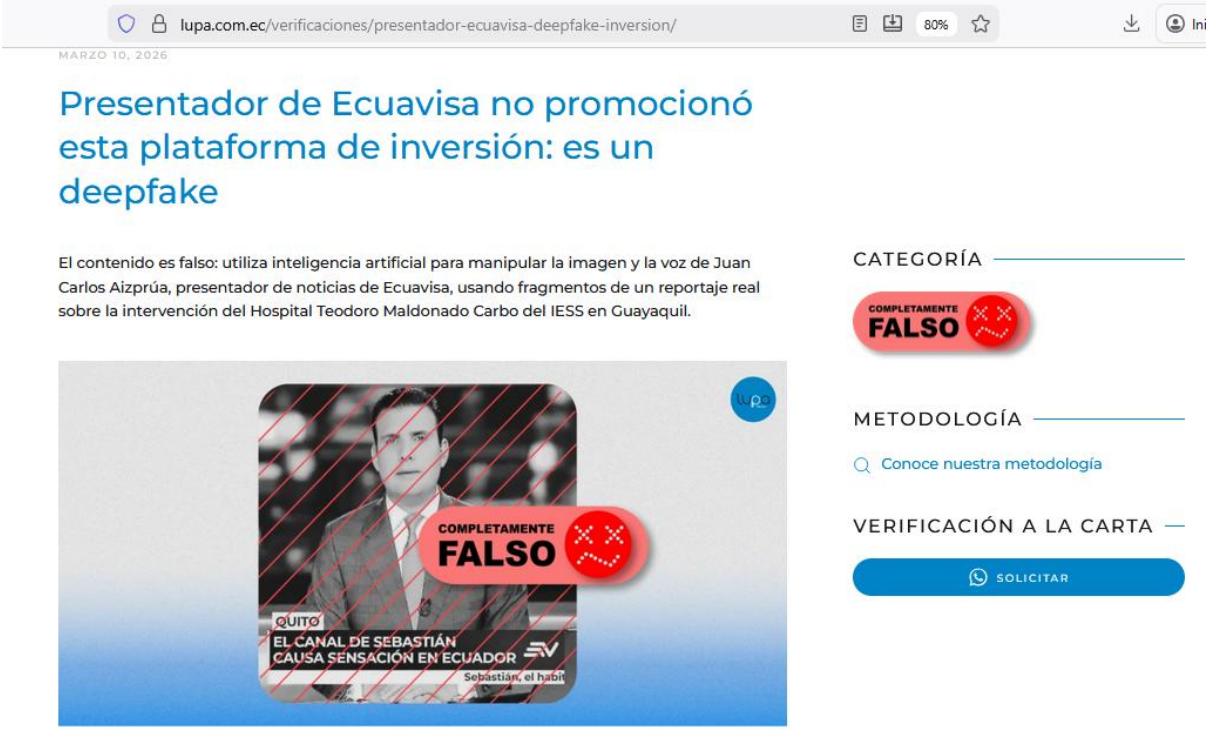
En marzo de 2026, la plataforma de verificación Lupa Media documentó la circulación de un video deepfake en redes sociales que suplantaba la imagen y la voz de un presentador de noticias del canal Ecuavisa, con el fin de promocionar una supuesta plataforma de inversión con rendimientos mínimos. El contenido utilizó fragmentos de un reportaje legítimo en el que el periodista informaba sobre la intervención en un hospital en Guayaquil, y los alteró mediante inteligencia artificial para fabricar un anuncio financiero fraudulento.

La cuenta que difundía el contenido contaba con al menos seis anuncios activos en Instagram y Facebook, redirigía a usuarios de Telegram identificados como bots y registraba ubicación en Hong Kong, con datos de contacto inactivos.

Este caso ilustra la secuencia estructurada descrita en el presente boletín: recolección de material audiovisual público, generación de contenido falso mediante inteligencia artificial, difusión masiva a través de plataformas digitales con pauta pagada y captación de víctimas mediante promesas de ganancias rápidas (Lupa Media, 2026).

Figura 2

Deepfake real relacionado a suplantación de identidad



Nota: *captura de pantalla caso real de deepfake* ***Fuente:*** *Lupa Media*

Medidas de Prevención y Mitigación

El uso de deepfakes en estafas digitales representa una amenaza para la seguridad digital y para la confianza en los entornos de interacción virtual. En Ecuador, este fenómeno demanda la adopción de medidas integrales de prevención y mitigación que articulen respuestas normativas, técnicas e interinstitucionales, con participación activa tanto de las instituciones del Estado como de la ciudadanía.

Como medida de contención, distintas plataformas digitales han adoptado políticas para restringir este tipo de contenidos. Facebook, por ejemplo, prohibió desde enero de 2020 determinados deepfakes, con excepción de aquellos claramente identificables como parodias. De igual manera, empresas tecnológicas como Twitter y Google han impulsado herramientas orientadas a detectar y limitar la difusión de contenidos multimedia falsificados (INCIBE, 2020).

Ante este escenario, la prevención requiere medidas que permitan reducir el riesgo y fortalecer la capacidad de respuesta de la ciudadanía, instituciones y plataformas tecnológicas, entre las cuales se destacan:

Medidas de prevención para la ciudadanía

Para la ciudadanía, la prevención se sustenta en la verificación de identidad, la protección de la información personal y la actuación oportuna ante situaciones sospechosas. En este marco, se destacan las siguientes medidas:

- Verificar la fuente de los contenidos multimedia sensibles, especialmente aquellos que generan reacción social.
- Evitar compartir material cuya autenticidad no haya sido contrastada por medios confiables o fuentes oficiales.
- Confirmar toda solicitud inesperada de dinero, información sensible o decisiones urgentes mediante un segundo canal de contacto, como una llamada directa o un medio alternativo de confianza.
- Aplicar preguntas de validación personal o mecanismos previamente acordados que permitan confirmar la identidad del interlocutor.
- Evitar decisiones inmediatas en escenarios de presión, considerando que este tipo de estafas suele apoyarse en la urgencia para reducir la verificación.
- Proteger la huella digital, limitando la publicación de imágenes de carácter íntimo, sexual o estrictamente personal, reforzando la seguridad de redes sociales y evitando el envío de material sensible incluso en entornos aparentemente privados.
- Conocer los derechos contemplados en la Ley Orgánica de Protección de Datos Personales (acceso, rectificación, eliminación) y solicitar la eliminación de contenidos manipulados que afecten la imagen o los datos personales.
- Denunciar de manera inmediata ante la Fiscalía General del Estado o la Policía Nacional cualquier uso de deepfakes en casos de extorsión, secuestro, sextorsión, fraude o difamación, preservando la evidencia digital, enlaces, mensajes, audios, imágenes o números de contacto.

Coordinación institucional e interinstitucional

La respuesta frente a este fenómeno requiere la articulación entre entidades públicas, operadores de justicia y actores tecnológicos. En este marco, la coordinación se traduce en acciones como las siguientes:

- Protocolos unificados para recepción, análisis y trámite adecuado de denuncias relacionadas con deepfakes, diferenciando casos de violencia digital, fraude, desinformación o afectación a autoridades e infraestructura.
- Programas de capacitación continua a fiscales, jueces, investigadores y peritos en técnicas especializadas para detección de deepfakes, preservación de evidencia digital y presentación clara y oportuna de las pruebas en los procesos judiciales.

- Campañas de comunicación y sensibilización articuladas que transmitan en lenguaje sencillo, definiciones básicas sobre deepfakes, señales de manipulación y rutas de atención ante posibles vulneraciones de derechos.
- Integración de contenidos sobre deepfakes en los programas curriculares del sistema educativo nacional.
- Medidas de protección para víctimas, especialmente cuando el deepfake se vincula a violencia de género, acoso o riesgos a la integridad personal.

Capacidades técnicas para detección y análisis

Desde el punto de vista técnico, la respuesta frente a deepfakes requiere capacidades de detección y análisis forense apoyadas en investigaciones y peritajes especializados en inteligencia artificial y análisis de medios digitales. Es fundamental que los CSIRT sectoriales, el ECUCERT y las áreas de TI de entidades públicas y privadas integren en sus procesos de gestión de incidentes escenarios de suplantación de identidad, campañas de desinformación y manipulación audiovisual.

Uso de herramientas de detección de deepfakes

La detección de deepfakes constituye un componente complementario en la mitigación de estafas digitales, especialmente debido al alto nivel de realismo que pueden alcanzar estos contenidos.

En determinados casos pueden identificarse señales de posible manipulación, para lo cual existen herramientas especializadas como Deepware Scanner, Reality Defender o Microsoft Video Authenticator, que permiten analizar:

- Alteraciones en patrones de imagen y sonido.
- Metadatos del archivo digital.
- Señales biométricas inconsistentes.
- Comparación con contenido original previamente disponible.

Ante la sospecha de manipulación, resulta recomendable evitar la difusión del contenido, activar los mecanismos de reporte en la plataforma digital o institución correspondiente y ponerlo en conocimiento de las autoridades competentes, conforme a la Guía de Contenido Inapropiado emitida por el Ministerio del Interior (Ministerio del Interior, 2025).

Conclusiones

- Las estafas con deepfakes constituyen una evolución del fraude digital, al integrar inteligencia artificial y técnicas de ingeniería social dentro de una secuencia planificada y altamente adaptable, orientada a la suplantación de identidad y a la obtención de beneficios ilícitos.
- El uso de la voz y la imagen como elementos de autenticidad convierte a los deepfakes en un vector especialmente eficaz de ciberdelincuencia, ya que desplaza el engaño desde la infraestructura tecnológica hacia la confianza depositada en la identidad digital de las personas.
- En Ecuador, el fenómeno de las estafas digitales evidencia una transformación más que una disminución. Si bien las noticias del delito relacionadas con suplantación de identidad presentan una reducción moderada, las estafas muestran una tendencia creciente, lo que sugiere una migración hacia modalidades más sofisticadas apoyadas en inteligencia artificial y deepfakes.
- El impacto de los deepfakes trasciende la dimensión económica derivada de las estafas, al generar también efectos sociales, reputacionales e institucionales que debilitan la confianza en las comunicaciones digitales, en las entidades públicas, en el sistema financiero y en la integridad de la información que circula en el ecosistema digital ecuatoriano.
- Los deepfakes representan una amenaza creciente para la seguridad digital en Ecuador, al combinar recursos tecnológicos avanzados con manipulación psicológica. Frente a ello, se requiere una respuesta integral que articule prevención ciudadana, fortalecimiento institucional, capacidades investigativas especializadas y cooperación con actores tecnológicos, a fin de mitigar riesgos y proteger la confianza en el entorno digital.

Recomendaciones

- Fortalecer las capacidades institucionales de análisis e investigación sobre las estafas con deepfakes, incorporando criterios técnicos, jurídicos y operativos que permitan su identificación temprana, adecuada documentación y efectiva persecución penal.
- Impulsar acciones de sensibilización y prevención orientadas a la verificación de comunicaciones audiovisuales, especialmente en contextos en los que se soliciten transferencias económicas, validaciones urgentes o entrega de información sensible a través de canales digitales.
- Promover la implementación de protocolos de verificación de identidad en entornos digitales, especialmente en instituciones públicas, empresas y entidades financieras,

a fin de evitar que decisiones críticas dependan exclusivamente de audios, videos o mensajes recibidos por medios digitales.

- Fortalecer la coordinación interinstitucional entre las entidades competentes del Estado, incluidas la Policía Nacional, la Fiscalía General del Estado, los organismos de regulación y control, y las unidades especializadas en ciberseguridad, con el fin de articular la respuesta frente a incidentes relacionados con deepfakes, suplantación de identidad y fraude digital.

Glosario de Términos

Deepfakes (Falsificaciones Profundas): son archivos de vídeo, imagen o voz manipulados mediante un software de inteligencia artificial de modo que parezcan originales, auténticos y reales. El término “Deepfakes” combina la palabra “fake” (falso) y la palabra “deep”, proveniente de “deep learning” (aprendizaje profundo, que es un tipo de aprendizaje automático de la inteligencia artificial) (LISA INSTITUTE, 2026).

Deepvoice (Deepfakes de voz): son audios generados o modificados utilizando IA para imitar voces humanas de manera tan realista que es casi imposible distinguirlas de las originales (Beamud, 2024).

Evidencia Digital (Evidencia Electrónica): se conoce como evidencia digital al conjunto de datos en formato digital que pueden utilizarse en un tribunal para esclarecer los hechos de un delito. Algunas de estas evidencias pueden ser archivos con contenido, metadatos, conexiones de tráfico en la red, discos duros, tarjetas o memorias USB (Universidad Europea, 2024).

Dividendo mentiroso: fenómeno por el cual la existencia de deepfakes reales facilita que personas o actores públicos nieguen la autenticidad de evidencia legítima, alegando que se trata de contenido manipulado o falso, eludiendo así la rendición de cuentas sobre hechos que sí ocurrieron. Este efecto se intensifica a medida que crece el conocimiento público sobre los deepfakes, ya que aumenta la plausibilidad de tales negaciones (Chesney & Citron, 2019)

Modelos generativos: es un modelo de machine learning diseñado para crear nuevos datos similares a sus datos de entrenamiento. Los modelos generativos de inteligencia artificial (IA) aprenden los patrones y distribuciones de los datos de entrenamiento y, a continuación, aplican esos conocimientos para generar contenidos novedosos en respuesta a los nuevos datos de entrada (Belcic, 2025).

Dato sintético: son datos artificiales diseñados para imitar los datos reales. Se genera a través de métodos estadísticos o mediante el uso de técnicas de inteligencia artificial (IA) como el deep learning y la IA generativa. Pese a ser generados artificialmente, los datos sintéticos conservan las propiedades estadísticas subyacentes de los datos originales en los que se basan. Como tales, los conjuntos de datos sintéticos pueden complementar o incluso sustituir a los conjuntos de datos reales (Caballar, 2025).

BIBLIOGRAFÍA

Asobanca. (2025). *Law Journal.* Departamento Legal de ASOBANCA. Obtenido de https://asobanca.org.ec/wp-content/uploads/2025/10/Law-Journal_VF.pdf

Beamud, R. (17 de marzo de 2024). *Deepfake y Deep Voice: Entendiendo el fenómeno de la manipulación audiovisual.* Obtenido de https://www.escudodigital.com/escudo-tv/deepfakes-deep-voice-cuando-evidencia-audiovisual-se-desmonta.html

Belcic, I. (noviembre de 2025). *¿Qué es un modelo generativo? .* Obtenido de https://www.ibm.com/es-es/think/topics/generative-model

Caballar, R. D. (27 de noviembre de 2025). *¿Qué son los datos sintéticos?* . Obtenido de https://www.ibm.com/es-es/think/topics/synthetic-data

Chesney, B., & Citron, D. (diciembre de 2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review, 107*. Obtenido de https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security

Citron, D., & Chesney , R. (2019). Deepfakes and the new disinformation war. *Foreign Affairs*, 147-155. Obtenido de https://scholarship.law.bu.edu/shorter_works/76

Europol. (marzo de 2022). *Facing reality? Law enforcement and the challenge of deepfakes.* Observatory report from the Europol Innovation Lab. Luxembourg: Publications Office of the European Union. Obtenido de https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes

IBM. (2025). *Cost of a Data Breach, Report 2025.* Obtenido de www.ibm.com: https://www.ibm.com/downloads/documents/us-en/131cf87b20b31c91

INCIBE. (abril de 2020). *Deepfakes, ¿cómo se aprovechan de esta tecnología para engañarnos?* Obtenido de https://www.incibe.es/ciudadania/blog/deepfakes-como-se-aprovechan-de-esta-tecnologia-para-enganarnos

INCIBE. (8 de 08 de 2023). *Intento de fraude amoroso utilizando técnicas de deepfake para suplantar a un personaje público*. Obtenido de https://www.incibe.es/linea-de-ayuda-

en-ciberseguridad/casos-reales/intento-de-fraude-amoroso-utilizando-tecnicas-de-deepfake-para-suplantar-un-personaje-publico

INEC. (julio de 2025). *Tecnologías de la Información y Comunicación-TIC*. Obtenido de https://www.ecuadorencifras.gob.ec/tecnologias-de-la-informacion-y-comunicacion-tic/

Kaspersky. (24 de julio de 2025). *Ciberamenazas que imitan herramientas de inteligencia artificial, como ChatGPT, aumentan 115% en 2025*. Obtenido de https://latam.kaspersky.com/about/press-releases/ciberamenazas-que-imitan-herramientas-de-inteligencia-artificial-como-chatgpt-aumentan-115-en-2025-kaspersky

Kietzmann, J., McCarthy, I., Lee, L., & Kietzmann, T. (2020). Deepfakes: Trick or treat? *Business Horizons*, 135-146. Obtenido de https://doi.org/10.1016/j.bushor.2019.11.006

LISA INSTITUTE. (24 de marzo de 2026). *Deepfakes: Qué es, tipos, riesgos y amenazas*. Obtenido de https://www.lisainstitute.com/blogs/blog/deepfakes-tipos-consejos-riesgos-amenazas?_pos=1&_sid=f29717141&_ss=r

Lupa Media. (10 de marzo de 2026). *Presentador de Ecuavisa no promocionó esta plataforma de inversión: es un deepfake.* Obtenido de lupa.com.ec/verificaciones/presentador-ecuavisa-deepfake-inversion

Maldonado, C., Sánchez, A., Ramirez, D., & Hallo, K. (1 de diciembre de 2024). *Revista Dilemas contemporaneos.* Obtenido de https://doi.org/10.46377/dilemas.v12i.4491

Mentinno. (2025). *Ecuador Estado Digital.* Recuperado el 2025, de Mentinno: https://www.mentinno.com/informesdigitales

Ministerio del Interior. (2025). *Ministerio del Interior.* Obtenido de https://www.ministeriodelinterior.gob.ec/wp-content/plugins/download-monitor/download.php?id=2050&force=0

Miranda Romero, A. (2025). Deepfake como estrategia para desinformar en las redes sociales durante las campañas electorales en Ecuador. *Revista de Ciencias Sociales, 31*(4), 223-238.

Romero N., W., Sinaluisa S., F., & Freire , N. (Julio-diciembre de 2024). Deepfakes Pornográficos: Impacto jurídico-probatorio y social en el Ecuador. *Reincisol, 3*(6), 2912-2934. Obtenido de https://doi.org/10.59282/reincisol.V3(6)2912-2934

Universidad Europea. (23 de febrero de 2024). *¿Qué es la evidencia digital? Características y tipos.* Obtenido de https://universidadeuropea.com/blog/evidencia-digital/

Universidad Politécnica Salesiana. (2025). La manipulación de la evidencia de políticas públicas con la Inteligencia Artificial Generativa: los riesgos de los deepfakes. *Universitas, 43*.

Vaccari, C., & Chadwick, A. (2020). *Deepfakes and disinformation: Exploring the impact of synthetic political video.* New Media & Society. Obtenido de https://doi.org/10.1177/2056305120903408

WatchGuard. (12 de junio de 2025). *La psicología del engaño o cómo protegerse del fraude digital*. Obtenido de https://www.watchguard.com/es/wgrd-news/blog/la-psicologia-del-engano-o-como-protegerse-del-fraude-digital-0