REVISTA INVECOM
ISSN en línea: 2739-0063

Extorsión digital con inteligencia artificial en el Perú: revisión sistemática del marco normativo y lineamientos para su prevención

Digital extortion using artificial intelligence in Peru: a systematic review of the regulatory framework and guidelines for its prevention

Recibido: 20/08/2025 - **Aceptado:** 17/11/2025

Sarasara Victoria Palomino Firma
https://orcid.org/0000-0002-1001-4380
spalominofi@ucvvirtual.edu.pe
Universidad César Vallejo. Lima, Perú

Resumen

Este artículo tiene como finalidad examinar el marco normativo peruano sobre extorsión digital potenciada por IA, compararlo con estándares internacionales y proponer lineamientos normativo-operativos para su prevención, persecución y prueba pericial. Empleando metodología de revisión sistemática (PRISMA 2020) de 30 fuentes académicas, normativas e institucionales (2020–2025), con análisis documental, jurídico-comparado y síntesis cualitativa temática, obteniéndose como resultados que, la evidencia muestra la capacidad de generación sintética (deepfakes y clonación de voz) supera la detección humana y algorítmica en contextos reales; el Perú avanzó con la Ley N.º 32314 (agravante por uso de IA), pero persisten vacíos de procedimientos de retiro ágil de contenidos, preservación temprana de evidencia y estándares periciales para audio/imagen sintética. En el derecho comparado, la UE (AI Act), Ofcom (Reino Unido) y eSafety (Australia) establecen obligaciones de diligencia, transparencia y herramientas de atribución/provenance. En consecuencia, se recomienda la implementación de un modelo híbrido complemente la sanción penal con deberes de plataformas (detección, etiquetado y preservación), fortalecimiento de laboratorios forenses y programas de prevención y educación en ciberseguridad, para mejorar la respuesta estatal y la protección de las víctimas.

Palabras clave: extorsión digital, inteligencia artificial, deepfakes

Abstract

This article aims to examine the Peruvian regulatory framework on AI-enhanced digital extortion, compare it with international standards, and propose normative–operational guidelines for its prevention, prosecution, and forensic examination. Using a systematic review methodology (PRISMA 2020) of 30 academic, regulatory, and institutional sources (2020–2025), with documentary, legal-comparative, and thematic qualitative analysis, the findings show that synthetic-content generation capabilities (deepfakes and voice cloning) surpass human and algorithmic detection in real-world contexts. Peru has made progress with Law No. 32314 (aggravating circumstance for the use of AI); however, gaps remain in procedures for the swift removal of harmful content, early preservation of evidence, and forensic standards for synthetic audio/image analysis. In comparative law, the EU (AI Act), Ofcom (United Kingdom), and eSafety (Australia) establish obligations relating to due diligence, transparency, and attribution/provenance tools. Consequently, this study concludes by recommending a hybrid model that complements criminal sanctions with platform duties (detection, labeling, and preservation), strengthening of forensic laboratories, and cybersecurity prevention and education programs, in order to improve the State's response and enhance victim protection.

Keywords: digital extortion, artificial intelligence, deepfakes

Introducción

El delito de extorsión ha experimentado una evolución significativa en las últimas dos décadas, pasando de modalidades presenciales tradicionales a esquemas digitales altamente sofisticados. Esta transformación se ha visto impulsada por la masificación de internet, el anonimato en redes sociales y el desarrollo de la inteligencia artificial (IA), que ha permitido nuevas formas de suplantación de identidad, manipulación audiovisual y chantaje en entornos virtuales (Europol, 2025; Holt et al., 2022).

En el Perú, la extorsión digital se ha convertido en una de las amenazas emergentes más preocupantes dentro del panorama de la cibercriminalidad. Según el Centro Nacional de Seguridad Digital (CNSD), entre 2023 y 2025 los reportes de sextorsión, extorsión a empresas y amenazas digitales aumentaron en un 78 %, coincidiendo con la proliferación de herramientas de clonación de voz y generación de deepfakes. Estas técnicas han sido utilizadas para: (i) suplantar la voz o imagen de familiares en simulaciones de secuestro virtual, (ii) extorsionar a menores y adolescentes mediante imágenes íntimas manipuladas (sextorsión), y (iii) amenazar a empresas y autoridades con filtraciones falsas, utilizando audios o videos generados por IA.

Los deepfakes, definidos como contenidos sintéticos hiperrealistas generados mediante redes generativas adversarias (GAN), constituyen hoy uno de los instrumentos predilectos del cibercrimen para la extorsión digital. Su facilidad de producción y viralización incrementa la vulnerabilidad de individuos y organizaciones (Mirsky y Lee, 2021; Regan, 2025). En varios países se han documentado casos de pérdidas millonarias por fraudes y chantajes derivados de esta tecnología, incluyendo videollamadas manipuladas en tiempo real que engañaron a directivos de multinacionales (Guembe et al., 2022).

En respuesta a estas amenazas, el Perú promulgó la Ley N° 32314 en abril de 2025, que modifica el Código Penal y la Ley de Delitos Informáticos (Ley 30096) para agravar la responsabilidad penal cuando un delito es cometido mediante inteligencia artificial. La norma incluye delitos como: (i) extorsión digital y sextorsión, (ii) suplantación de identidad con audios o videos generados por IA, y (iii) fraudes y chantajes corporativos basados en deepfakes (Maldonado, 2025; Proyecto Zero 24, 2025).

No obstante, persisten retos importantes, tales como la ausencia de jurisprudencia específica, lo que dificulta la aplicación uniforme de la agravante, la limitada capacidad pericial, debido a la carencia de laboratorios especializados en detección de deepfakes y análisis forense de IA, y la falta de coordinación con plataformas digitales, que en otros países están obligadas a remover contenido ilícito en plazos breves (European Commission, 2025; UK Government, 2025).

A nivel internacional, el tratamiento de la extorsión digital con IA evidencia tres tendencias:

1. **Punitiva:** Estados Unidos aplica la Computer Fraud and Abuse Act (CFAA) y leyes estatales que sancionan la difusión de deepfakes maliciosos y la sextorsión digital (Holt et al., 2022; (Shetty, et al., 2024).
2. **Preventiva:** La Unión Europea, a través del Artificial Intelligence Act (2024) y el Digital Services Act (DSA), exige etiquetado obligatorio de contenido sintético y retiro inmediato de deepfakes que constituyan delitos (European Parliament y Council, 2024)
3. **Mixta:** Reino Unido y Australia combinan sanciones penales con responsabilidad de plataformas para prevenir la propagación de contenido ilícito (eSafety Commissioner, 2025, 2023; UK Government, 2025).

En Latinoamérica, la regulación es incipiente, podemos señalar los países de México y Colombia solo contemplan agravantes tecnológicas sin protocolos de remoción de contenido ni tipificación autónoma de deepfakes (CNDH, 2025; Inter-American Development Bank y Organization of American States, 2020).

Mientras que, en el Perú, aunque avanzó con la Ley 32314, carece de mecanismos preventivos integrales, quedando rezagado frente a modelos híbridos anglosajones y europeos.

El objetivo general del presente artículo es analizar el marco normativo peruano sobre la extorsión digital con IA, compararlo con estándares internacionales y proponer lineamientos normativo-operativos para su prevención, persecución y prueba pericial.

Metodología

Se llevó a cabo un estudio cualitativo, documental y comparativo con el propósito de analizar e interpretar el marco normativo peruano sobre la extorsión digital potenciada por inteligencia artificial, comparándolo con estándares internacionales. El diseño metodológico fue hermenéutico-interpretativo, empleado para la exégesis normativa y doctrinal, y complementado con un enfoque fenomenológico para comprender el fenómeno en su contexto sociojurídico. Esta aproximación sigue las líneas metodológicas habituales en trabajos de revisión, basándose en un enfoque cualitativo, una revisión bibliográfica rigurosa, criterios específicos de selección y exclusión, y un relato detallado del proceso de búsqueda.

Este enfoque está respaldado por manuales reconocidos de investigación cualitativa revisados sistemáticamente en las ciencias sociales y el derecho, tales como Hernández-Sampieri et al. (2018), Flick (2015) y Kitchenham y Charters (2007). Para la planificación, cribado y reporte de la revisión sistemática se aplicó el estándar PRISMA 2020, manteniendo un registro interno exhaustivo del protocolo que incluye objetivos, fuentes consultadas, periodo de análisis, criterios de selección y una matriz de extracción de datos. El reporte final se ajusta a la checklist PRISMA e incluye el diagrama de flujo de la revisión (véase la Figura PRISMA del manuscrito) (Page et al., 2021).

La búsqueda se realizó para el período 2020-2025 y abarcó diversas bases de datos, incluyendo Scopus, Web of Science, ScienceDirect, SpringerLink, SciELO, Redalyc y Google Scholar. Además, se incluyó literatura gris y normativa relevante. Se emplearon descriptores en ambos idiomas, español e inglés, utilizando operadores booleanos que enfatizan términos relacionados con extorsión digital, inteligencia artificial y aspectos normativos vinculados tanto a Perú como a estándares internacionales.

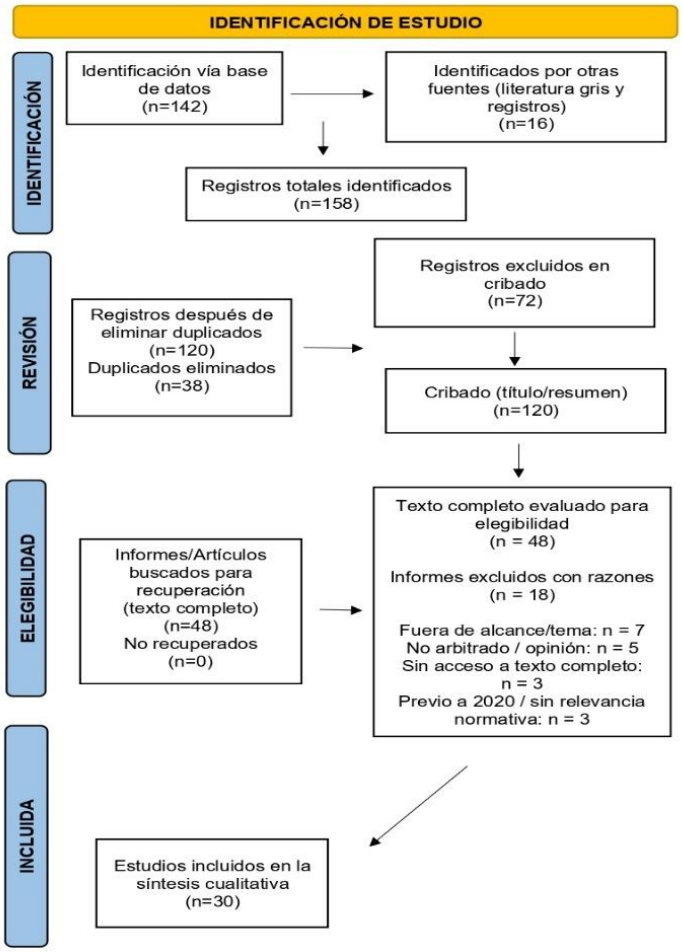
Para complementar esta estrategia, se realizaron exploraciones adicionales en portales oficiales de Perú, como el Congreso y la Presidencia del Consejo de Ministros (PCM). Asimismo, se consultaron fuentes regulatorias y normativas de otras regiones, incluyendo la Unión Europea, Reino Unido y Australia, con el objetivo de recolectar normativa primaria y guías regulatorias pertinentes al tema.

Los criterios de inclusión abarcaron artículos arbitrados del período indicado, normativas y documentos oficiales relevantes a la extorsión digital, inteligencia artificial, deepfakes, sextorsión y responsabilidad de plataformas, en español o inglés con acceso a texto completo. Se excluyeron duplicados, artículos de opinión sin revisión por pares, documentos anteriores a 2020 salvo normas seminales, y textos no relacionados directamente con extorsión o inteligencia artificial.

El proceso de selección conforme a PRISMA se inició con 158 registros identificados, depurando hasta 120 tras cribado y eliminación de duplicados, y seleccionando 48 para evaluación de texto completo. Finalmente, 30 documentos fueron incluidos en la síntesis cualitativa, entre ellos artículos científicos, normativos e informes oficiales, reflejados en el diagrama PRISMA adjunto.

Figura 1

Diagrama de flujo PRISMA



Para la extracción y evaluación de la calidad, se construyó una matriz que demostró varios aspectos fundamentales, entre ellos el tipo de fuente, la jurisdicción aplicable, la tipificación del delito, las preventivas existentes, el papel que desempeñan las plataformas digitales, las evidencias forenses disponibles y los vacíos legales identificados en la legislación vigente.

En la valoración de los artículos arbitrados se aplicaron listas PRISMA y criterios cualitativos específicos que permitieron evaluar la claridad de los objetivos planteados, la coherencia de la metodología utilizada y la validez de los resultados obtenidos, asegurando un análisis riguroso y sistemático.

Por otro lado, en el análisis de la normativa oficial, se priorizó el texto consolidado y los documentos emitidos por autoridades reconocidas. Entre ellos se destacan la AI Act y la Digital Services Act (DSA) disponibles en EUR-Lex, así como el Online Safety Act en GOV.UK, que sirven de referencia para la regulación internacional en materia de inteligencia artificial y seguridad digital.

La síntesis y el análisis de los datos se llevaron a cabo mediante técnicas de análisis temático y hermenéutico. Estos permitieron identificar patrones de respuesta tanto punitivos como preventivos y mixtos, y también detectar brechas regulatorias como la ausencia de protocolos para la rápida remoción de contenidos ilícitos o la falta de peritajes especializados en deepfakes y clonación de voz.

Además, se evidenciaron oportunidades para mejorar la alineación normativa con las obligaciones que deben cumplir las plataformas digitales, lo cual es clave para una gestión efectiva de la extorsión digital y otros delitos cibernéticos relacionados.

Se elaboró una matriz comparativa que confrontó la situación normativa del Perú con la de jurisdicciones internacionales representativas, incluyendo la Unión Europea, Estados Unidos, Reino Unido, Australia y países de América Latina y el Caribe, integrando tanto los hallazgos legales como la literatura académica especializada.

Respecto a las consideraciones éticas, el estudio se fundamentó exclusivamente en el uso de fuentes públicas y no involucró la participación de sujetos humanos, por lo que no fue necesaria la aprobación de un comité de ética.

Finalmente, se reconocieron ciertas limitaciones del estudio, como la heterogeneidad de las fuentes consultadas, el sesgo potencial debido a la disponibilidad documental, y la naturaleza dinámica de la regulación legal, que pudo haber experimentado cambios posteriores al cierre de la búsqueda. Para afrontar estas limitaciones, se priorizó la consulta de fuentes oficiales confiables y se aplicaron criterios claros de reproducibilidad utilizando el estándar PRISMA.

Resultados y discusión

La revisión integró 30 fuentes organizadas en cuatro bloques: (i) técnico-forense (creación/detección de deepfakes y clonación de voz); (ii) victimización y sextorsión; (iii) análisis jurídico-regulatorio; y (iv) tendencias operativas/institucionales. La calidad metodológica fue adecuada para una síntesis cualitativa y permitió triangulación entre evidencia técnica, social y normativa.

Los estudios de estado del arte y revisiones muestran un avance acelerado en modelos generativos (GAN, TTS/VC) y herramientas de manipulación multimodal, a la par de defensas todavía frágiles ante compresión, ruido y cambios de canal (Mirsky y Lee, 2021; Rana et al., 2022; Tolosana et al., 2020; Verdoliva, 2020). En audio, las encuestas sistemáticas documentan brechas entre resultados de laboratorio y desempeño “en la vida real”, con débil generalización de detectores frente a ataques fuera de distribución (Khanjani et al., 2023; Yi et al., 2023; Zhang et al., 2025). Los retos ASVspoof 2021 reportan mejoras, pero confirman la vulnerabilidad de los detectores al migrar a escenarios “in-the-wild” (Yamagishi et al., 2021; Liu et al., 2022). Trabajos recientes exploran biomarcadores acústicos y ataques parcialmente falsos capaces de potenciar fraudes tipo “CEO” y extorsión por voz clonada (Alali y Theodorakopoulos, 2025; Kulangareth et al., 2024). En síntesis, el bloque técnico justifica protocolos periciales robustos con bancos de prueba, validación cruzada y reporte explícito de incertidumbre (Rana et al., 2022; Yi et al., 2023).

La literatura identifica una expansión sostenida de la sextorsión y del material íntimo sintético no consensual (NSII), con impacto desproporcionado en mujeres, adolescentes y grupos vulnerables (Laffier, 2023; Ray et al., 2024). La detección humana de manipulación audiovisual es limitada y heterogénea entre modalidades (audio/imagen/video), lo que facilita el engaño y el chantaje (Diel et al., 2024). Estudios policiales y cualitativos perfilan patrones de contacto-coerción-escalamiento y obstáculos probatorios (Dolev-Cohen y Levy, 2022; Edwards et al., 2023), mientras que análisis comparativos en 10 países evidencian prevalencia y actitudes que normalizan NSII (Umbach et al., 2024). Factores emocionales y de personalidad aumentan la exposición al riesgo (Tzani et al., 2024).

El panorama regulatorio es desigual: la UE aprobó el AI Act con obligaciones de transparencia y gestión de riesgo (European Parliament y Council, 2024); el Reino Unido (Ofcom) y Australia (eSafety) publicaron guías operativas y toolkits de atribución para industria y plataformas (eSafety Commissioner, 2025; Ofcom, 2024; Ofcom, 2025). Para el Perú, la Ley N° 32314 introduce la agravante por uso de IA en delitos como la extorsión, pero persisten vacíos procedimentales de retiro ágil, preservación de evidencia y estándares periciales (Perú –

Congreso, 2025). La doctrina y los estudios comparados convergen en la necesidad de modelos híbridos que combinen sanción penal, deberes de diligencia de plataformas y medidas de provenance/etiquetado (Guerrero-Sierra y Díaz, 2025; Sandoval y Sharman, 2024; Kira, 2024; Ma'arif y Rahman, 2025; McGlynn et al., 2025; Shirish y Komal, 2024; Sulaj, 2024).

A nivel operativo, el IOCTA 2025 documenta extorsión multimodal (vishing con voz clonada + doxxing + difusión) y la profesionalización de servicios criminales basados en IA y datos robados, reforzando la urgencia de cooperación transfronteriza y atribución técnica (Europol, 2025).

Los resultados evidencian una asimetría entre la capacidad ofensiva de generación sintética y la capacidad defensiva de detección: los detectores aún no generalizan con fiabilidad fuera del laboratorio (Alali y Theodorakopoulos, 2025; Khanjani et al., 2023; Kulangareth et al., 2024; Liu et al., 2022; Mirsky y Lee, 2021; Rana et al., 2022Verdoliva, 2020; Yamagishi et al., 2021; Yi et al., 2023; Zhang et al., 2025). Sumado al bajo desempeño humano en identificar deepfakes (Diel et al., 2024) y al contexto de NSII/sextorsión (Dolev-Cohen y Levy, 2022; Edwards et al., 2023; Laffier, 2023; Ray et al., 2024; Tzani et al., 2024; Umbach et al., 2024), se conforma un entorno propicio para la extorsión digital.

La Ley N° 32314 aporta disuasión punitiva, pero carece de procedimientos operativos equivalentes a los modelos híbridos comparados: takedown en plazos breves, preservación temprana y etiquetado/provenance del contenido sintético (eSafety Commissioner, 2025; European Parliament y Council, 2024; Ofcom, 2024, 2025; Perú – Congreso, 2025). La evidencia sugiere formalizar protocolos periciales (cadena de custodia para audio/imagen, uso de conjuntos de referencia tipo ASVspoof y reporte de incertidumbre) y vías rápidas para víctimas, especialmente en casos de sextorsión (Guerrero-Sierra y Díaz, 2025; Kira, 2024; Ma'arif y Rahman, 2025; McGlynn et al., 2025; Sandoval y Sharman, 2024; Shirish y Komal, 2024; Sulaj, 2024; Europol, 2025).

Cabe mencionar que, como recomendaciones normativo-operativas, mencionamos lo siguiente: (i) Peritaje y prueba digital: Incorporar bancos de prueba y validaciones "in-the-wild" (ASVspoof), con métricas estandarizadas y reporte de incertidumbre en dictámenes (Liu et al., 2022; Rana et al., 2022; Yamagishi et al., 2021; Yi et al., 2023). (ii) Deberes de plataformas: Establecer plazos perentorios de retiro, órdenes de preservación y marcadores de procedencia/etiquetado (European Parliament y Council, 2024; eSafety Commissioner, 2025; Ofcom, 2024, 2025;). (iii) Protección de víctimas: Canales de denuncia y contención (plantillas de retiro, asistencia psicosocial) y verificación fuera de banda para cortar el vishing con voz clonada (Dolev-Cohen y Levy, 2022; Edwards et al., 2023; Laffier, 2023; Ray et al., 2024; Tzani et al., 2024;). (iv) Cooperación internacional: Mecanismos automáticos de asistencia mutua y atribución en línea con IOCTA 2025 (Europol, 2025).

La síntesis depende de estudios con heterogeneidad metodológica y de un marco regulatorio en rápida evolución; se mitiga con triangulación y selección de fuentes primarias. Futuros trabajos deberían evaluar efectividad real de detectores y modelos de gobernanza (provenance, marcados, transparencia) mediante pilotos sectoriales y auditorías periódicas (Ma'arif y Rahman, 2025; McGlynn et al., 2025; Rana et al., 2022; Yi et al., 2023).

Conclusiones

La extorsión digital basada en IA es una amenaza presente y escalable en el Perú. La evidencia revisada (2020–2025) muestra que la madurez técnica de los deepfakes y la clonación de voz supera la capacidad humana y algorítmica de detección en contextos "reales", creando un entorno propicio para la extorsión, el fraude y la sextorsión. La respuesta estatal debe asumir que el riesgo es estructural y creciente.

El enfoque punitivo peruano es necesario, pero insuficiente por sí solo. La Ley N° 32314 introduce la agravante por uso de IA y envía una señal disuasiva relevante; sin embargo, persisten brechas operativas: ausencia de protocolos de retiro ágil (takedown), preservación temprana de evidencia digital, estándares periciales para audio/imágenes sintéticas y coordinación interinstitucional.

Se impone un modelo híbrido regulatorio-operativo. La comparación internacional sugiere combinar sanción penal con deberes de diligencia para plataformas (detección, etiquetado de contenido sintético y conservación de datos), herramientas de atribución/proveniencia y plazos perentorios de atención a órdenes judiciales y administrativas. Este enfoque eleva la probabilidad de disrupción temprana y éxito probatorio.

Asimismo, la capacidad pericial, la investigación y la prueba requieren laboratorios con bancos de prueba y evaluaciones "in-the-wild" (p. ej., retos tipo ASVspoof), métricas robustas (AUC, EER por canal/compresión), reporte de incertidumbre y cadenas de custodia digital específicas para contenido sintético. Sin esta infraestructura, la persecución penal pierde eficacia.

En cuanto a la protección integral de víctimas y prevención primaria, la prevalencia de NSII/sextorsión demanda rutas de denuncia simplificadas, órdenes de preservación inmediatas a plataformas, acompañamiento

psicosocial y alfabetización digital (p. ej., verificación “fuera de banda” y códigos seguros familiares) en escuelas, universidades y lugares de trabajo.

La cooperación transfronteriza y gobernanza de datos: la extorsión digital cruza jurisdicciones y se apalanca en servicios globales; por ello, se requieren canales de asistencia mutua estandarizados, intercambio de indicadores técnicos (hashes, firmas, marcas de agua) y participación en redes internacionales para atribución técnica y remoción coordinada.

Por otro lado, sobre las limitaciones y agenda futura, la rápida evolución tecnológica y regulatoria puede volver obsoletos algunos hallazgos; además, la heterogeneidad metodológica de la literatura limita inferencias causales. Se recomienda: (a) pilotos sectoriales de etiquetado/proveniencia con medición de impacto; (b) auditorías periódicas de detectores en escenarios reales; (c) estudios sobre la costo-efectividad de protocolos de preservación; y (d) evaluación de resultados en víctimas tras implementar rutas de retiro y contención.

Referencias

Alali, A., y Theodorakopoulos, G. (2025). Partial fake speech attacks in the real world using deepfake audio. *Journal of Cybersecurity and Privacy, 5*(1), 6. https://doi.org/10.3390/jcp5010006

Chawki, M. (2024). Navigating legal challenges of deepfakes in the American landscape. *Cogent Social Sciences, 10*(1). https://doi.org/10.1080/23311816.2024.2320971

CNDH. (enero de 2025). Informe Anual 2024. Comisión Nacional de los Derechos Humanos. https://www.cndh.org.mx/sites/default/files/documentos/2025-01/InformeAnual_2024.pdf

Congreso de la República Perú . (29 de April de 2025). *Ley N° 32314 (modifica el Código Penal y la Ley 30096 para incluir el uso de IA en la comisión de delitos)*. https://leyes.congreso.gob.pe/Documentos/2021_2026/ADLP/Texto_Consolidado/32314-TXM.pdf

Diel, A., Freetly, A., y Walther, E. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis. *Heliyon Computer Science*. https://doi.org/10.1155/hbe2/1833228

Dolev-Cohen, M., y Levy, N. (2022). A qualitative examination of school counselors’ perspectives of online sextortion. *International Journal of Bullying Prevention, 4*(4), 293-309. https://doi.org/10.1007/s42380-022-00159-0

Edwards, M., Pitts, M., y O’Malley, R. (2023). Online sextortion: Characteristics of offences from a policing perspective. *Journal of Policing and Society*. https://doi.org/10.1016/j.jps.2023.100038

eSafety Commissioner. (28 de mayo de 2025). Deepfake trends and challenges — position statement. https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes

eSafety Commissioner. (27 de June de 2025). *Guide to responding to image-based abuse involving AI deepfakes*. https://www.esafety.gov.au/sites/default/files/2025-06/Respond%203A%20-%20Guide%20to%20responding%20to%20image-based%20abuse%20involving%20AI%20deepfakes.pdf

European Commission. (4 de febrero de 2025). Guidelines on prohibited AI practices (AI Act). https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

European Parliament, y Council. (12 de july de 2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139.* Official Journal of the European. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Europol. (2025). *Internet Organised Crime Threat Assessment.* La Haya: Europol: IOCTA 2025. https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2025

Flick, U. (2015). *Introducción a la investigación cualitativa* (5 ed.). Morata. https://study.sagepub.com/flick_aiqr

Guembe, B., Azeta, A., Sanjay , M., Chukwudi Osamor, V., Fernández-Sanz, L., y Pospelova, V. (2022). La amenaza emergente de los ciberataques impulsados por IA: un análisis. *Taylor&Francis Group, 36*(1), 2409. https://doi.org/10.1080/08839514.2022.2037254

Guerrero-Sierra, H., y Díaz, L. (2025). Threats and regulatory challenges of non-consensual deepfakes. *Cogent Social Sciences, 11*(1). https://doi.org/10.1080/23311886.2025.2552342

Hernández-Sampieri, R., y Mendoza, C. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta.* McGraw-Hill Educación. https://doi.org/10.22201/fesc.20072236e.2019.10.18.6

Holt, T. J., Bossler, A. M., y Seigfried-Spellar, K. C. (2022). *Cybercrime and Digital Forensics* (3ra ed.). Routledge. https://www.routledge.com/Cybercrime-and-Digital-Forensics-An-Introduction/Holt-Bossler-Seigfried-Spellar/p/book/9780367360078?srsltid=AfmBOoqE8Ppwy8iZFMBH2HbQu4fNHTOj6zJ706cz_6xa-m9a7MhfXblU

Inter-American Development Bank, y Organization of American States. (2020). 2020 Cybersecurity Report: Risks, Progress, and the Way Forward in Latin America and the Caribbean. *IDB*. http://dx.doi.org/10.18235/0002513

Khanjani, Z., Watson, G., y Janeja, V. (2023). Audio deepfakes: A survey. Frontiers in Big Data. *Frontiers in Bits Data*. https://doi.org/10.3389/fdata.2022.1001063

Kira, B. (2024). When non-consensual intimate deepfakes go viral: Legal responses and platform duties. *Computer Law & Security Review, 53*. https://doi.org/10.1016/j.clsr.2024.105033

Kitchenham, B., y Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. EBSE Technical Report, Keele University. https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf

Kulangareth, N., Nath, A., Davy, J., y Mantri, A. (2024). Investigation of deepfake voice detection using speech biomarkers. *JMIR Biomedical Engineering, 9*(1). https://doi.org/10.2196/56245

Liu, X., Wang, X., Sahidullah, M., Patino, J., Delgado, H., Kinnunen, T., y Nautsch, A. (2022). ASVspoof 2021: Towards spoofed and deepfake speech detection in the wild. *arXiv preprint arXiv:2210.02437, 31*, 2507-2522. https://doi.org/10.1109/TASLP.2023.3285283

Ma'arif, A., y Rahman, M. (2025). Social, legal, and ethical implications of AI-generated deepfakes. *Journal of Responsible Technology*. https://doi.org/10.1016/j.jrt.2025.100610

Maldonado, V. (29 de abril de 2025). *Ley 32314: Modifican el Código Penal y otro para sancionar el uso de la inteligencia artificial en la comisión de delitos.* LP Pasión por el derecho: https://lpderecho.pe/ley-32314-modifican-codigo-penal-sancionar-uso-inteligencia-artificial-delitos/

McGlynn, C., Rackley, E., y Johnson, K. (2025). The "new voyeurism": Criminalizing the creation of deepfake porn. *Journal of Law and Society, 52*(2), 215-240. https://doi.org/10.1111/jols.12527

Mirsky, Y., y Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys, 54*(1), 1-41. https://doi.org/10.1145/3425780

Ofcom. (23 de July de 2024). *Deepfake defences: Mitigating the harms of deceptive synthetic media*. https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/deepfake-defences

Ofcom. (11 de July de 2025). *Deepfake Defences 2 – The Attribution Toolkit.* . https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/deepfake-defences-2--the-attribution-toolkit

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., y Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, n*(71), 372. https://doi.org/10.1136/bmj.n71

Proyecto Zero, 2. (29 de abril de 2025). *Proyecto Zero 24.* Boletín Legal PZ-24: Ley N.º 32314. Proyecto Zero 24: https://proyectozero24.com/wp-content/uploads/2025/04/Boletin-Legal-PZ-24-2025-LEY-N%C2%B0-32314.pdf

Rana, M., Nobi, M., Murali, B., y Sung, A. (2022). Deepfake detection: A systematic literature review. *IEEE Access, 10*, 25494-25513. https://doi.org/10.1109/ACCESS.2022.3154404

Ray., A., Henry, N., y Flynn, A. (2024). Sextortion: A scoping review. *Trauma, Violence, & Abuse (online first)*. https://doi.org/10.1177/15248380241277271

Regan, G. (18 de junio de 2025). The State of Deepfake Regulations in 2025: What Businesses. *Reality Defender*. https://www.realitydefender.com/insights/the-state-of-deepfake-regulations-in-2025-what-businesses-need-to-know

Sandoval, M., y Sharman, J. (2024). Threat of deepfakes to the criminal justice system: A systematic review. *Crime Science, 13*(2). https://doi.org/10.1186/s40163-024-00239-1

Shetty, S., Choi, K., y Park, I. (2024). Investigating the Intersection of AI and Cybercrime: Risks, Trends, and Countermeasures. *International Journal of Cybersecurity Intelligence & Cybercrime, 7*(2). https://vc.bridgew.edu/ijcic/vol7/iss2/3

Shirish, A., y Komal, A. (2024). A socio-legal inquiry on deepfakes. *California Western International Law Journal, 55*(1), 1-36. https://scholarlycommons.law.cwsl.edu/cwilj/vol55/iss1/3

Sulaj, A. (2024). Legal responses to challenges posed by deepfakes. *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5146075

Tolosana, R., Vera-Rodríguez, R., Fierrez, J., Morales , A., y Ortega-García, J. (2020). DeepFakes and beyond: A survey of face manipulation and fake detection. *Information Fusion, 64*, 131-148. https://doi.org/10.1016/j.inffus.2020.06.014

Tzani, C., Grigoriou, C., y Panagiotopoulos, C. (2024). The role of personality, emotional factors and sexual needs in sextortion. *Computers in Human Behavior, 152*. https://doi.org/10.1016/j.chb.2024.108177

UK Government. (24 de abril de 2025). Guidance Online Safety Act: explainer. https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer

Umbach, R., Henry, N., Beard, G., y Berryessa, C. (2024). Non-consensual synthetic intimate imagery: Prevalence, attitudes, and knowledge in 10 countries. *CHI '24 Proceedings*. https://doi.org/10.1145/3613904.3642382

Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing, 14*(5), 910–932. https://doi.org/10.1109/JSTSP.2020.3002101

Yamagishi, J., Wang, X., Todisco, M., Sahidullah, M., Patino, J., Nautsch, A., y Evans, N. (2021). ASVspoof 2021: Accelerating progress in spoofed and deepfake speech detection. *arXiv preprint arXiv:2109.00537*. https://arxiv.org/abs/2109.00537

Yi, J., Wang, C., Tao, J., Zhang, X., Zhang, C., y Zhao, Y. (2023). Audio deepfake detection: A survey. *arXiv preprint arXiv:2308.14970.* https://arxiv.org/abs/2308.14970

Zhang, B., Lin, Z., Yi, J., Wu, J., y Xu, X. (2025). Audio Deepfake Detection: What Has Been Achieved and What Lies Ahead. *Sensors MPDI, 25*(7), 1989. https://doi.org/10.3390/s25071989