REVISTA INVECOM
ISSN en línea: 2739-0063

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

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

Received: 20/08/2025 - **Accepted:** 17/11/2025

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

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

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

Introduction

The crime of extortion has undergone significant evolution over the past two decades, transitioning from traditional in-person methods to highly sophisticated digital schemes. This transformation has been driven by the widespread proliferation of the internet, anonymity on social media, and advancements in artificial intelligence (AI), which have facilitated new forms of identity theft, audio-visual manipulation, and virtual blackmail (Europol, 2025; Holt et al., 2022).

SciELO Venezuela latindex LatinREV CC BY

In Peru, digital extortion has emerged as one of the most concerning threats within the landscape of cybercrime. According to the National Center for Digital Security (CNSD), reports of sextortion, extortion against businesses, and digital threats increased by 78% between 2023 and 2025, coinciding with the rise of voice cloning tools and deepfake generation. These techniques have been employed to: (i) impersonate the voice or image of family members in simulations of virtual kidnapping, (ii) extort minors and adolescents using manipulated intimate images (sextortion), and (iii) threaten businesses and authorities with false leak claims, utilizing audios or videos generated by AI.

Deepfakes, defined as hyper-realistic synthetic content produced using generative adversarial networks (GANs), have become one of the preferred instruments of cybercrime for digital extortion. Their ease of production and potential for virality heighten the vulnerability of individuals and organizations (Mirsky and Lee, 2021; Regan, 2025). Numerous countries have documented cases of multimillion-dollar losses resulting from fraud and blackmail stemming from this technology, including manipulated video calls in real time that deceived executives of multinational corporations (Guembe et al., 2022).

In response to these threats, Peru enacted Law No. 32314 in April 2025, which amends the Penal Code and the Computer Crimes Law (Law 30096) to intensify criminal liability when a crime is committed using artificial intelligence. This statute addresses offenses such as: (i) digital extortion and sextortion, (ii) identity theft involving audios or videos generated by AI, and (iii) corporate fraud and blackmail based on deepfakes (Maldonado, 2025; Proyecto Zero 24, 2025).

Despite these advancements, significant challenges persist, including the lack of specific jurisprudence, which hampers the uniform application of the aggravating circumstances, limited expert capacity due to the absence of specialized laboratories for deepfake detection and AI forensic analysis, and the lack of coordination with digital platforms, which in other countries are mandated to remove illegal content within short time frames (European Commission, 2025; UK Government, 2025).

Internationally, the treatment of digital extortion involving AI reveals three trends:

1. **Punitive:** The United States enforces the Computer Fraud and Abuse Act (CFAA) and state laws that penalize the dissemination of malicious deepfakes and digital sextortion (Holt et al., 2022; Shetty et al., 2024).
2. **Preventive:** The European Union, through the Artificial Intelligence Act (2024) and the Digital Services Act (DSA), mandates compulsory labeling of synthetic content and immediate removal of deepfakes that constitute criminal offenses (European Parliament & Council, 2024)
3. **Mixed:** The United Kingdom and Australia combine penal sanctions with platform liability to prevent the spread of illegal content (eSafety Commissioner, 2025; UK Government, 2025).

In Latin America, regulation is still in its infancy; countries such as Mexico and Colombia only contemplate technological aggravating factors without protocols for content removal or autonomous classification of deepfakes (CNDH, 2025; Inter-American Development Bank and Organization of American States, 2020).

In Peru, while progress has been made with Law 32314, there remains a lack of comprehensive preventive mechanisms, leaving it lagging behind hybrid Anglo-Saxon and European models.

The overarching objective of this article is to analyze Peru's regulatory framework concerning digital extortion involving AI, compare it with international standards, and propose normative-operational guidelines for its prevention, prosecution, and forensic examination.

Methodology

A qualitative, documentary, and comparative study was conducted to analyze and interpret Peru's regulatory framework regarding digital extortion fueled by artificial intelligence, juxtaposing it with international standards. The methodological design was hermeneutic-interpretive, employed for normative and doctrinal exegesis, and complemented by a phenomenological approach to understand the phenomenon within its socio-legal context. This approach adheres to the conventional methodological lines in review work, relying on a qualitative focus, rigorous bibliographic review, specific selection and exclusion criteria, and a detailed narrative of the search process.

This methodology is supported by recognized manuals for qualitative research systematically reviewed in social sciences and law, including Hernández-Sampieri et al. (2018), Flick (2015), and Kitchenham and Charters (2007). For the planning, screening, and reporting of the systematic review, the PRISMA 2020 standard was applied, maintaining a comprehensive internal record of the protocol that includes objectives, consulted sources, analysis period, selection criteria, and a data extraction matrix. The final report adheres to the PRISMA checklist and includes the flow diagram of the review (see the PRISMA figure in the manuscript) (Page et al., 2021).

The search covered the period from 2020 to 2025 and encompassed various databases, including Scopus, Web of Science, ScienceDirect, SpringerLink, SciELO, Redalyc, and Google Scholar. Additionally, relevant gray

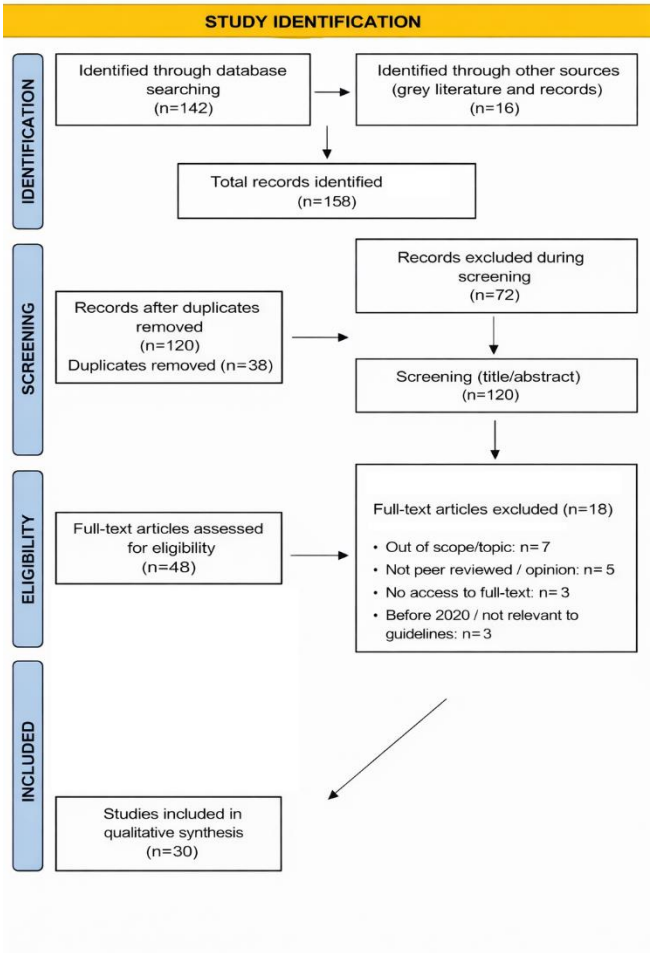
literature and regulations were included. Descriptors in both Spanish and English were employed, utilizing boolean operators that highlighted terms related to digital extortion, artificial intelligence, and regulatory aspects pertinent to both Peru and international standards.

To complement this strategy, additional explorations were conducted on official Peruvian portals, such as the Congress and the Presidency of the Council of Ministers (PCM). Furthermore, regulatory and normative sources from other regions, including the European Union, the United Kingdom, and Australia, were consulted to gather relevant primary legislation and regulatory guidelines on the subject.

Inclusion criteria encompassed peer-reviewed articles from the specified period, regulations, and official documents relevant to digital extortion, artificial intelligence, deepfakes, sextortion, and platform liability, in either Spanish or English with full text access. Duplicates, opinion pieces without peer review, documents predating 2020 except seminal regulations, and texts not directly related to extortion or artificial intelligence were excluded.

The selection process according to PRISMA initiated with 158 identified records, narrowed down to 120 after screening and duplicates removal, and 48 were selected for full-text evaluation. Ultimately, 30 documents were included in the qualitative synthesis, consisting of scientific articles, regulatory texts, and official reports, as reflected in the attached PRISMA diagram.

Figure 1
PRISMA flow diagram



For the extraction and evaluation of quality, a matrix was constructed to demonstrate various fundamental aspects, including the type of source, applicable jurisdiction, crime classification, existing preventive measures, the role of digital platforms, available forensic evidence, and identified legal gaps within current legislation.

In assessing the peer-reviewed articles, PRISMA checklists and specific qualitative criteria were applied, allowing for the evaluation of the clarity of stated objectives, coherence of the employed methodology, and validity of the obtained results, ensuring a rigorous and systematic analysis.

Furthermore, in the analysis of official regulations, priority was given to consolidated texts and documents issued by recognized authorities. Key documents include the AI Act and the Digital Services Act (DSA) available on EUR-Lex, as well as the Online Safety Act on GOV.UK, which serve as benchmarks for international regulation concerning artificial intelligence and digital security.

The synthesis and analysis of data were conducted using thematic and hermeneutic analysis techniques. These approaches enabled the identification of response patterns that are punitive, preventive, and mixed, while also revealing regulatory gaps such as the absence of protocols for the rapid removal of illegal content and the lack of specialized expertise in deepfakes and voice cloning.

Moreover, opportunities for improving regulatory alignment with the obligations of digital platforms were identified, which is crucial for effective management of digital extortion and related cybercrimes.

A comparative matrix was developed, juxtaposing the regulatory situation in Peru with that of representative international jurisdictions, including the European Union, the United States, the United Kingdom, Australia, and countries in Latin America and the Caribbean. This matrix integrated both legal findings and specialized academic literature.

Regarding ethical considerations, the study was based exclusively on publicly available sources and did not involve human subjects, negating the need for ethical committee approval.

Finally, several limitations of the study were recognized, such as the heterogeneity of consulted sources, potential bias due to the availability of documents, and the dynamic nature of legal regulation, which may have experienced changes after the search period concluded. To address these limitations, reliance was prioritized on trustworthy official sources, and clear reproducibility criteria were applied, following the PRISMA standard.

Results and discussion

The review incorporated 30 sources organized into four categories: (i) technical-forensic (creation/detection of deepfakes and voice cloning); (ii) victimization and sextortion; (iii) legal-regulatory analysis; and (iv) operational/institutional trends. The methodological quality was adequate for a qualitative synthesis, allowing for triangulation between technical, social, and normative evidence.

State-of-the-art studies and reviews reveal rapid advancements in generative models (GAN, TTS/VC) and multimodal manipulation tools, alongside still-fragile defenses against compression, noise, and channel changes (Mirsky & Lee, 2021; Rana et al., 2022; Tolosana et al., 2020; Verdoliva, 2020). In audio, systematic surveys document gaps between laboratory results and real-world performance, with weak generalization of detectors against out-of-distribution attacks (Khanjani et al., 2023; Yi et al., 2023; Zhang et al., 2025). The ASVspoof 2021 challenges report improvements but confirm the vulnerability of detectors when migrating to “in-the-wild” scenarios (Yamagishi et al., 2021; Liu et al., 2022). Recent studies explore acoustic biomarkers and partially false attacks that can augment “CEO”-type fraud and voice cloning extortion (Alali & Theodorakopoulos, 2025; Kulangareth et al., 2024). In summary, the technical block justifies robust expert protocols with test banks, cross-validation, and explicit uncertainty reporting (Rana et al., 2022; Yi et al., 2023).

The literature identifies a sustained expansion of sextortion and non-consensual intimate synthetic material (NSII), disproportionately impacting women, adolescents, and vulnerable groups (Laffier, 2023; Ray et al., 2024). Human detection of audio-visual manipulation is limited and heterogeneous across modalities (audio/image/video), facilitating deception and blackmail (Diel et al., 2024). Police studies and qualitative research outline patterns of contact-coercion-escalation and evidentiary obstacles (Dolev-Cohen and Levy, 2022; Edwards et al., 2023), while comparative analyses across ten countries highlight the prevalence and attitudes that normalize NSII (Umbach et al., 2024). Emotional and personality factors increase exposure to risk (Tzani et al., 2024).

The regulatory landscape is uneven: the EU has approved the AI Act with transparency and risk management obligations (European Parliament & Council, 2024); the United Kingdom (Ofcom) and Australia (eSafety) have published operational guidelines and attribution toolkits for industry and platforms (eSafety Commissioner, 2025; Ofcom, 2024; Ofcom, 2025). For Peru, Law No. 32314 introduces aggravating circumstances for the use of AI in crimes such as extortion; however, procedural gaps remain regarding agile removal, evidence preservation, and expert standards (Peru – Congress, 2025). The doctrine and comparative studies converge on the necessity of hybrid models that combine penal sanctions, platform diligence obligations, and provenance/labeling measures (Guerrero-Sierra & Díaz, 2025; Sandoval & Sharman, 2024; Kira, 2024; Ma’arif & Rahman, 2025; McGlynn et al., 2025; Shirish & Komal, 2024; Sulaj, 2024).

Operationally, the IOCTA 2025 documents multimodal extortion (vishing with cloned voice + doxxing + dissemination) and the professionalization of criminal services based on AI and stolen data, underscoring the urgent need for transnational cooperation and technical attribution (Europol, 2025).

The findings reveal an asymmetry between the offensive capacity for synthetic generation and the defensive capacity for detection: detectors still do not generalize reliably outside the laboratory (Alali & Theodorakopoulos, 2025; Khanjani et al., 2023; Kulangareth et al., 2024; Liu et al., 2022; Mirsky & Lee, 2021; Rana et al., 2022; Verdoliva, 2020; Yamagishi et al., 2021; Yi et al., 2023; Zhang et al., 2025). Coupled with low human performance in identifying deepfakes (Diel et al., 2024) and the context of NSII/sextortion (Dolev-Cohen & Levy, 2022; Edwards et al., 2023; Laffier, 2023; Ray et al., 2024; Tzani et al., 2024; Umbach et al., 2024), a conducive environment for digital extortion is established.

Law No. 32314 offers punitive deterrence; however, it lacks operational procedures equivalent to those of the compared hybrid models: prompt takedowns, early preservation, and labeling/provenance of synthetic content (eSafety Commissioner, 2025; European Parliament & Council, 2024; Ofcom, 2024, 2025; Peru – Congress, 2025). Evidence suggests formalizing expert protocols (chain of custody for audio/image, use of ASVspoof-type reference sets, and uncertainty reporting) and rapid response pathways for victims, particularly in sextortion cases (Guerrero-Sierra & Díaz, 2025; Kira, 2024; Ma'arif & Rahman, 2025; McGlynn et al., 2025; Sandoval & Sharman, 2024; Shirish & Komal, 2024; Sulaj, 2024; Europol, 2025).

As normative-operational recommendations, we propose the following: (i) Forensic and digital evidence: Incorporate test banks and "in-the-wild" validations (ASVspoof), with standardized metrics and uncertainty reporting in assessments (Liu et al., 2022; Rana et al., 2022; Yamagishi et al., 2021; Yi et al., 2023). (ii) Platform obligations: Establish mandatory deadlines for takedown, preservation orders, and markers of provenance/labeling (European Parliament & Council, 2024; eSafety Commissioner, 2025; Ofcom, 2024, 2025). (iii) Victim protection: Reporting and containment channels (removal templates, psychosocial support) and out-of-band verification to disrupt vishing with cloned voice (Dolev-Cohen & Levy, 2022; Edwards et al., 2023; Laffier, 2023; Ray et al., 2024; Tzani et al., 2024). (iv) International cooperation: Automatic mechanisms for mutual assistance and online attribution in accordance with IOCTA 2025 (Europol, 2025).

The synthesis depends on studies with methodological heterogeneity and a rapidly evolving regulatory framework; it is mitigated by triangulation and the selection of primary sources. Future work should assess the real effectiveness of detectors and governance models (provenance, labeling, transparency) through sectorial pilots and periodic audits (Ma'arif & Rahman, 2025; McGlynn et al., 2025; Rana et al., 2022; Yi et al., 2023).

Conclusions

Digital extortion based on artificial intelligence (AI) represents a present and scalable threat in Peru. The reviewed evidence (2020–2025) indicates that the technical maturity of deepfakes and voice cloning surpasses the human and algorithmic capacity for detection in "real-world" contexts, creating a conducive environment for extortion, fraud, and sextortion. The state response must recognize that this risk is both structural and increasing.

While the Peruvian punitive approach is necessary, it is insufficient on its own. Law No. 32314 introduces an aggravating factor for the use of AI and sends an important deterrent signal; however, operational gaps remain, including the lack of agile takedown protocols, early preservation of digital evidence, expert standards for synthetic audio/images, and inter-institutional coordination.

A hybrid regulatory-operational model is essential. International comparisons suggest that penal sanctions should be combined with diligence obligations for platforms (detection, labeling of synthetic content, and data preservation), attribution/provenance tools, and strict deadlines for responding to judicial and administrative orders. This approach increases the likelihood of early disruption and successful prosecution.

Moreover, forensic capacity, research, and evidence collection necessitate laboratories equipped with test banks and "in-the-wild" evaluations (e.g., ASVspoof-type challenges), robust metrics (AUC, EER by channel/compression), uncertainty reporting, and specific digital chains of custody for synthetic content. Without this infrastructure, criminal prosecution loses efficacy.

Regarding comprehensive victim protection and primary prevention, the prevalence of non-consensual intimate synthetic material (NSII) and sextortion demands simplified reporting channels, immediate preservation orders to platforms, psychosocial support, and digital literacy (e.g., out-of-band verification and secure family codes) in schools, universities, and workplaces.

Cross-border cooperation and data governance are critical, as digital extortion transcends jurisdictions and leverages global services. Standardized mutual assistance channels, the exchange of technical indicators (hashes, signatures, watermarks), and participation in international networks for technical attribution and coordinated removal are necessary.

On the topic of limitations and future agenda, rapid technological and regulatory evolution can render some findings obsolete, and the methodological heterogeneity of the literature limits causal inferences. Recommendations include: (a) sector-based pilot programs for labeling/provenance with impact measurement; (b)

periodic audits of detectors in real-world scenarios; (c) studies on the cost-effectiveness of preservation protocols; and (d) evaluation of victim outcomes after implementing removal and containment pathways.

References

Alali, A., & 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., & 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., & 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., & 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, & 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., & 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., & 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., & 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., & 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, & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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., & 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