Sci

MDPI

Review

A Review of Crime at Machine Speed: Criminological Aspects of Artificial Intelligence's Industrialisation of Deception

Paolo Bailo [1], Ascanio Sirignano [1,*], Giulio Nittari [2], Giuseppe Visconti [1], Giuliano Pesel [1], Tommaso Spasari [3] and Giovanna Ricci [1]

[1] Section of Legal Medicine, School of Law, University of Camerino, 62032 Camerino, Italy; paolo.bailo@unicam.it (P.B.); giuseppe.visconti@studenti.unicam.it (G.V.); dr.giuliano.pesel@gmail.com (G.P.); giovanna.ricci@unicam.it (G.R.)
[2] Telemedicine and Telepharmacy Centre, School of Medicinal and Health Products Sciences, University of Camerino, 62032 Camerino, Italy; giulio.nittari@unicam.it
[3] Forensic Medicine and BioLaw, Niccolò Cusano University, 00166 Rome, Italy
* Correspondence: ascanio.sirignano@unicam.it; Tel.: +39-0737-402435

Abstract

Artificial intelligence (AI) is transforming criminal practice by industrialising deception, compressing attack cycles, and corroding evidentiary trust. This narrative review synthesises recent technical and criminological literature with institutional reporting to explain how generative models, predictive analytics, and automation enable convincing synthetic media, highly targeted social engineering, document forgery, identity synthesis, and adaptive evasion. Attention is given to the convergence with organised networks that use AI to coordinate logistics, mimic normal behaviour, and launder proceeds across platforms. Furthermore, a review of the grey literature was carried out to identify applied cases and to show how heterogeneous they are. Defensive efforts are advancing, yet detection remains brittle under laundering, increasing media realism, and adversarial adaptation. Regulatory and policy responses are surveyed across jurisdictions without claiming exhaustiveness; they appear fragmented and often lag operational innovation. The objective is pragmatic: to raise attacker costs and preserve information integrity while safeguarding fundamental rights and forensic reliability.

Keywords: artificial intelligence; cybercrime; deepfakes; voice cloning; social engineering; adversarial machine learning

Academic Editors: Carson K. Leung and Haridimos Kondylakis

Received: 4 November 2025
Revised: 11 February 2026
Accepted: 24 February 2026
Published: 2 March 2026

1. Introduction

The advent of artificial intelligence (AI) constitutes a structural inflexion in the evolution of cybercrime, reconfiguring offenders' modus operandi and amplifying the sophistication, scale, and stealth of malicious activity [1,2]. Table 1 provides a concise chronology of these developments.

Over the past two decades, cybercrime has undergone a radical transformation from rudimentary intrusion methods to complex, automated operations. AI has catalysed and amplified this trajectory, rendering threats faster, more adaptive, and harder to detect. According to the EU Serious and Organised Crime Threat Assessment 2025 (EU-SOCTA 2025), AI is "accelerating organized crime," particularly through targeted digital fraud, deepfakes, voice cloning, and highly personalised phishing schemes [3]. Social engineering, identity theft, and digital money laundering are increasingly enhanced by algorithms capable of processing large datasets and simulating human-like behaviour. The increasing

convergence of criminal and state actors further complicates the global landscape, as attacks are motivated not only by profit but also by geopolitical disruption.

Table 1. AI–Cybercrime Milestones (2010–2025): Key Events and Security Impacts.

Year	Event	Impact on Cybersecurity
2010	Proliferation of machine learning	Early malware families began using machine-learning techniques to evade signature-based antivirus mechanisms.
2015	Integration of AI in defence systems	Organisations deploy AI systems for anomaly detection; adversaries adopt more advanced evasion tactics.
2018	Automated AI-driven attacks	Emergence of intelligent botnets and personalised phishing using NLP.
2020	Deepfakes and advanced social engineering	Generative AI produces realistic synthetic media (images, voices, video), increasing the believability and reach of social-engineering attacks.
2023	Generative AI in malware	Malware adapts to target environments in real time, making detection and mitigation significantly more difficult; adversarial-ML studies show evasion of static and dynamic detection via training-data poisoning and behaviour shifting that ML models struggle to generalise.
2025	Multimodal AI and predictive attacks	Attackers use predictive analytics and external data sources (web, social media, dark web) to anticipate defences and tailor attacks with greater precision; forecasting models based on web-forum signals report promising F1-score improvements for cyberattack prediction.

One of the most alarming developments is the emergence of adaptive malware, facilitated by adversarial machine learning. Patil et al. demonstrated that AI-based malware detection systems can be evaded through carefully crafted adversarial samples, unless models are explicitly designed for robustness [4]. Similarly, Chen et al. described how poisoning attacks can manipulate training data, undermining the reliability of AI-driven detection systems [5]. These dynamics underscore a co-adaptive threat environment in which defenders and offenders continuously learn and adapt, limiting the durability of static countermeasures.

AI is also leveraged to strengthen phishing and social engineering techniques. By analysing large datasets, generative models can produce highly personalised and credible messages, deceiving victims into disclosing sensitive information or performing harmful actions [6]. Empirical evidence indicates that such AI-assisted phishing campaigns achieve click-through rates comparable to or higher than human-generated attacks [7]. The evolution of phishing and social engineering has been accelerated by advances in machine learning and Natural Language Processing (NLP), enabling text that closely mimics authentic communication at scale. Jabir et al. systematically reviewed human factors in AI-powered phishing and found that cognitive biases, overconfidence, and automation trust significantly increase vulnerability to AI-generated deception [8].

Recent research also highlights the predictive potential of AI for both offensive and defensive cyber operations. Goyal et al. demonstrated that predictive analytics using web-based and dark web signals can forecast cyberattacks with notable accuracy [9], while Danish proposed real-time models that enhance proactive threat detection [10]. However, Alavizadeh et al. argue that predictive AI in cybersecurity must be conceptualised within a dynamic "Markov game" framework, where both attackers and defenders continuously learn and adapt [11]. AI is also used to automate large-scale attacks such as Distributed Denial of Service (DDoS), enabling real-time coordination that amplifies both effectiveness and complexity.

Institutional reports corroborate these findings. Europol (2025) emphasises that AI is enhancing organised crime operations across multiple domains—including drug trafficking, human trafficking, and money laundering—by automating and optimising criminal processes [3]. This intersection of automation, deception, and transnational coordination defines a new paradigm in AI-enabled criminality.

This article, through a narrative review and illustrative cases, provides an overview of how artificial intelligence can be leveraged to facilitate criminal undertakings, including impersonation via the generation of synthetic content (from voice deepfakes to full-body personae), the exploitation of financial resources through fictitious identities, and the operational support it can provide to organised crime. It also provides a concise analysis of regulatory measures adopted by governments against the criminal use of AI. Specifically, it examines how AI reshapes criminal opportunity structures and decision-making, and why countermeasures often fail under adversarial adaptation.

2. Materials and Methods

This study employs a documentary narrative review encompassing technical, criminological, and legal sources, selected to reflect the interdisciplinary and rapidly evolving nature of AI-enabled offending. Searches were conducted in major bibliographic databases—Scopus, Web of Science (WoS), IEEE Xplore, ACM Digital Library, PubMed, SSRN—and in the arXiv preprint server to capture both peer-reviewed and frontier research (January 2010–October 2025). Boolean strings combined AI capabilities with offence categories (e.g., "artificial intelligence" AND cybercrime; deepfake; "voice cloning" AND fraud; "synthetic identit*" AND laundering; "adversarial machine learning" AND evasion; "organized crime" AND AI). Coverage was expanded through backward/forward citation tracking and hand-searches of institutional repositories (Europol, INTERPOL, UNODC, ENISA, NIST, MITRE). Relevant statutory and regulatory materials were identified via official gazettes and consolidated legal databases.

Given that operational descriptions of emergent cyber-enabled crime frequently appear earlier in institutional and technical reporting than in peer-reviewed literature, a structured search of grey literature published between 2018 and October 2025 in English and Italian was also conducted. This included policy and regulatory reports, law-enforcement advisories, threat-intelligence briefs, white papers, technical blogs, and investigative journalism, and was executed via Google/Google Scholar and targeted domain queries using a predefined keyword set (e.g., "AI-enabled deception," "voice-cloning fraud," "automated phishing," "LLM-assisted social engineering"). Grey sources were used to document observed tactics, techniques, and incident archetypes rather than to infer prevalence or causal effects. Records were screened using the AACODS checklist; inclusion required concrete incident detail and triangulation across at least two independent sources, while duplicates, purely speculative pieces, and non-substantiated claims were excluded. This approach aligns with established expectations for narrative reviews, prioritising transparency of selection choices and explicit handling of evidence heterogeneity. To strengthen the methodological value of this narrative review, we conducted a SANRA (Scale for the Assessment of Narrative Review Articles) -informed appraisal of the narrative review articles cited as background evidence within the manuscript. The resulting item-level scores are reported in the Supplementary Materials to enhance transparency regarding the quality and evidentiary weight of secondary sources.

3. Deepfakes, Voice Cloning, and the Criminal Exploitation of Synthetic Media

The rise in AI has enabled the creation of highly realistic synthetic multimedia content—commonly referred to as deepfakes—that can be deployed for illicit purposes. Such manipulated media can depict real individuals in fabricated or compromising scenarios, with severe consequences for privacy, reputation, and psychological wellbeing [12]. Deepfakes exemplify one of the most ethically complex manifestations of the current AI revolution, challenging traditional notions of truth, authenticity, and human identity [13,14]. In criminological terms, synthetic media expands opportunity structures by lowering production costs, increasing plausibility, and compressing the time required to operationalise deception at scale, while simultaneously degrading the ordinary cues by which victims and institutions have historically assessed credibility.

Recent advances in generative adversarial networks (GANs) and diffusion models have dramatically reduced the technical barriers to producing synthetic video and audio, enabling their use in sextortion and non-consensual intimate imagery crimes [15]. In such cases, offenders create fake but photorealistic material that overlays a victim's face onto explicit scenes and threaten disclosure unless a ransom is paid. Even when no authentic intimate material exists, the fabricated content can appear genuine enough to destroy reputations or induce severe trauma [12]. Harms manifest at the individual level through acute psychological distress—humiliation, anxiety, fear, and loss of autonomy—alongside reputational and relational damage that jeopardises employability and personal safety, and generates concrete economic losses via extortion, identity fraud, and costly remediation [16,17]. At the institutional level, deepfakes increase caseload burdens for platforms, employers, and law enforcement while straining takedown pathways and data-protection remedies; at the systemic level, they erode informational trust and amplify gender-based and reputational vulnerabilities, with disproportionate impacts on minors and public figures [18]. The proliferation of non-consensual deepfake pornography exemplifies these dynamics, intensifying victimisation while exposing the limits of reactive content moderation and after-the-fact legal repair.

The rapid proliferation of deepfakes is paralleled by voice-cloning technologies that reproduce human speech with high fidelity using deep neural architectures such as Tacotron 2, WaveNet, and VALL-E, enabling capture of timbre, prosody, and emotional tone from only a few seconds of audio [19,20]. Criminal exploitation is already evident: synthetic voices are deployed for fraud, identity theft, and social engineering—often defeating voice authentication systems and enabling CEO-impersonation and "virtual kidnapping" scams [21–26]. Demographic and situational factors—age, gender, and listener awareness—modulate attack success, while user studies show that humans frequently fail to distinguish bona fide from AI-generated speech, with error rates exceeding 60% [21,22,24]. Although machine- and deep-learning–based detectors (e.g., Convolutional Neural Networks (CNNs), GANs) can flag synthetic speech, their effectiveness degrades in real-world acoustic conditions, under linguistic variation, and when attackers apply laundering or other obfuscation strategies [27–30]. Adversaries increasingly rely on perturbation, replay, laundering, and ultrasonic methods to evade detection [28,31–34], illustrating a co-adaptive environment in which defensive gains are repeatedly stress-tested and partially neutralised [35–37]. Beyond financial crime, voice cloning undermines biometric authentication, enables scalable social engineering, and directly jeopardises the integrity of audio evidence in forensic contexts; proposed countermeasures such as watermarking and masking show promise but remain largely experimental [38,39].

In parallel, the justice system is increasingly facing not only a technological challenge but also an epistemological one: a transition from "trust in visual evidence" to certified

traceability of origin and provenance [14]. Judges and forensic experts can no longer treat the "expert eye" as a sufficient safeguard [40]. A shift toward deep digital forensics is required—one that examines cross-layer consistency across devices, file systems, and server-side logs rather than relying on surface plausibility. Contemporary courtroom threats now include advanced evidentiary fabrication, including LLM (Large Language Model)-enabled generation of entire chat databases (e.g., WhatsApp, Telegram) that reproduce idiolect, abbreviations, and habitual errors; AI-assisted fabrication of e-mails that simulates not only message bodies but also plausible header structures and coherent metadata; and synthetic "scanned" documents that emulate scan artefacts (background noise, page skew, faded stamps) to produce an appearance of age and authenticity for PDFs generated moments earlier [41]. These developments raise practical questions regarding the allocation of the burden of proof for authenticity verification and the proposed use of "AI against AI" to detect non-human patterns in generated texts or metadata, while remaining embedded in a continuing arms race of adaptation [42].

Legal countermeasures against deepfakes increasingly blend criminalisation, platform-governance duties, and evidentiary safeguards. The United Kingdom has announced new standalone offences for creating sexually explicit deepfakes without consent—complementing the Online Safety Act 2023—thereby shifting liability onto perpetrators rather than victims of image-based abuse [43,44]. At the EU level, the Artificial Intelligence Act requires transparency and labelling for AI-generated or manipulated content (including deepfakes), and the Digital Services Act strengthens platform notice-and-action and transparency obligations—together creating a compliance baseline for provenance and moderation [45,46]. Outside Europe, China's "Deep Synthesis" rules mandate content labelling and provider due-diligence, while Australia criminalised the non-consensual sharing of sexually explicit deepfakes at the federal level and strengthened civil takedown powers through the Online Safety Act [47,48]. Brazil's Superior Electoral Court prohibited deepfakes in campaign communications and ordered clear AI disclosures, exemplifying targeted electoral protections [49]. In the United States, a patchwork of state measures combines political-deepfake rules and personality-right/voice-cloning protections (e.g., Tennessee's 2024 ELVIS Act), while many states continue to adopt disclosure or blackout-period rules for AI-altered political ads [50,51]. South Korea, confronting a surge of synthetic-sex-crime cases, has tightened its Special Act on the Punishment of Sexual Crimes to raise penalties and to criminalise viewing/possession of sexually explicit deepfakes [52]. To protect evidentiary integrity, legal frameworks are increasingly paired with technical standards and forensic practice: courts and regulators encourage provenance (e.g., C2PA Content Provenance and Authenticity) and admit state-of-the-art detection grounded in convolutional neural networks, spatio-temporal analysis, and multimodal fusion, while recent surveys and studies map both technical advances and real-world fragilities that legal systems must anticipate under adversarial adaptation [53–56].

4. Automated Financial and Identity Frauds via AI

AI now enables financial fraud and identity theft at scale by processing massive datasets to simulate credible financial behaviours, defeating traditional banking controls and anti-fraud systems [57,58]. Identity fraud increasingly relies on AI to collect, correlate, and manipulate personal information from public sources and data breaches to fabricate realistic synthetic identities capable of opening bank accounts, obtaining credit, or subscribing to digital services in the names of unaware third parties [58,59]. It is therefore important to distinguish synthetic identity fraud from conventional identity theft: rather than merely hijacking an existing victim's account, offenders assemble partially real and partially fabricated attributes into a new, plausible persona that can persist across

institutions, accumulate "legitimacy" over time, and scale across automated onboarding pipelines. Generative AI further enhances phishing and social-engineering operations by crafting tailored messages and mimicking authentic banking interfaces, luring victims into authorising payments, divulging PINs/OTPs (Personal Identification Number/One-Time Password), or installing remote-access malware; it can also forge high-fidelity documents that bypass Know-Your-Customer (KYC) checks [58,60]. These capabilities increase reach while compressing operational timelines, enabling rapid iteration on persuasive scripts and continuously adapting lures to the victim's context.

On the defensive side, widely deployed detection pipelines include machine-learning classifiers—random forest, isolation forest, neural networks—and anomaly-detection methods that achieve high accuracy against evolving fraud patterns [61–71]. Generative-AI techniques are dual-use: they enable synthetic-identity creation and complex scams, but they also improve detection via predictive analytics and advanced anomaly detection [64,70,72,73]. AI systems increasingly integrate social-media signals for identity verification, combining ensemble methods with behavioural biometrics to strengthen risk scoring [74–78]. Academic advances show that hybrid deep-learning models—combining RNNs (Recurrent Neural Network), Transformers, Autoencoders, and mixture-of-experts frameworks—can identify synthetic identities and anomalous transaction sequences with high efficacy [79]. However, operational performance is uneven outside controlled settings: models can degrade under distribution shift, data sparsity, adversarial manipulation, and cross-channel laundering strategies that fragment signals across platforms. "Identity deepfake" attacks against biometric authentication (e.g., synthetic voice or image inputs) also remain under-appreciated by both experts and the public, exposing modern systems to spoofing, escalation, and downstream account compromise [80]. In this co-adaptive setting, defensive gains are often temporary unless detection stacks are continuously validated against evolving attacker tactics and integrated with procedural friction for high-risk actions.

Lawmakers are countering automated financial and identity fraud enabled by artificial intelligence through risk-based regulation, platform duties, and strengthened identity and payments controls. The European Union's Artificial Intelligence Act mandates transparency and risk management for higher-risk systems, while the Payment Services Directive requires strong customer authentication to curb account-takeover and payment fraud [45,81]. The United Kingdom imposes duties to reduce fraudulent advertising online and introduces mandatory identity verification for company controllers to deter shell identities and false filings [44,82]. In the United States, the communications regulator clarified that calls using synthetic voices fall under the Telephone Consumer Protection Act, enabling enforcement against voice-cloned scam calls [83]. Beyond Europe and North America, China targets organised online fraud with a dedicated national statute, Singapore authorises binding disruption orders against scam infrastructure, and Australia strengthens customer due diligence and reporting through anti-money-laundering law [84–86].

Systemically, Europol warns that AI is automating scams against banks, fintech firms, and private users, threatening digital economic stability and complicating law-enforcement efforts [3]. These developments underscore that AI-enabled financial and identity fraud is not merely a technical problem: regulatory gaps, social trust, and institutional capacity are directly implicated. Effective responses must therefore pair multi-factor and behavioural authentication with continuously updated detection stacks, rigorous provenance and onboarding controls, and governance mechanisms that anticipate adversarial adaptation rather than treating fraud models as static solutions [57,58].

5. Technologically Assisted Organised Crime and the Integration of Artificial Intelligence

Organised crime is rapidly integrating AI into core operations, producing "technologically assisted deviance" that automates logistics, simulates identities, manipulates financial systems, and weaponizes disinformation to impede detection [3,87,88]. Criminal networks increasingly resemble quasi-corporate structures with dedicated units for cybercrime, AI-driven fraud, and digital social engineering, turning AI into a force multiplier via algorithmic decision-making and predictive analytics [87–89]. At the same time, it is analytically important to distinguish between AI as an efficiency amplifier and AI as a genuinely transformative organisational force. In many settings, AI primarily increases speed, scale, targeting precision, and operational resilience within familiar criminal business models; in others, it enables forms of coordination, impersonation, and adaptive optimisation that change how networks recruit, allocate roles, manage risk, and learn from enforcement pressure. This distinction clarifies why AI-enabled organised crime is not reducible to "more cybercrime," but can also reconfigure the structure and governance of illicit enterprises.

Applications span deepfakes, voice cloning, and AI-enabled phishing for financial fraud, stock manipulation, and influence operations [3,90–95]. In the logistical sphere, predictive systems that ingest real-time data anticipate law-enforcement interventions, map surveillance blind spots, and optimise smuggling routes for drugs and human trafficking using clustering and geospatial analytics [95]. Tactically, autonomous drones and AI-coordinated vehicles enable persistent surveillance and remote delivery of illicit goods in otherwise inaccessible environments—an evolution already documented in Mexico, Haiti, parts of Africa, and emerging European contexts [96,97]. Financially, synthetic identity fraud—constructing credible identities from fragments of real data—has surged, with advanced harvesting and correlation pipelines enabling pass-through of KYC checks and illicit access to credit [89,98]. In money laundering and market abuse, machine learning, neural networks, and anomaly-detection techniques are used to disguise flows and evade surveillance [63,68,99–103], while "crime as a service," deepfake-based fraud, and increasingly autonomous criminal networks are forecast to expand [90,92,104–106]. Crucially, offenders adapt to anti-money-laundering controls by mimicking normal patterns and exploiting model blind spots, degrading detector performance over time [101,107–113]. The net effect is a self-optimising ecosystem in which algorithms refine illegal strategies in response to environmental feedback, strengthening operational security and complicating attribution and disruption—an "automation of organized crime" that strains traditional investigative and regulatory frameworks [3,87,95].

Several jurisdictions have begun to adopt targeted legal countermeasures that explicitly address the ways organised networks weaponise artificial intelligence. Italy's Law 23 September 2025 No. 132 establishes a new offence of "illicit diffusion of content generated or altered with artificial intelligence" and aggravates penalties where offences (including certain market-abuse crimes) are committed through artificial-intelligence systems—thereby directly linking criminal liability to the use of such technologies [114]. China's Law Against Telecom and Online Fraud creates a dedicated framework against transnational scam syndicates, imposing preventative responsibilities on telecommunications, financial and online-service providers and enabling coordinated disruption of fraud infrastructures [84]. Singapore's Online Criminal Harms Act authorises binding directions—such as access-blocking, takedown and account-restriction orders—to disrupt scam infrastructure and impersonation campaigns at source, including before a substantive offence is complete [85,115]. The United Kingdom's Online Safety Act imposes duties on large online services to reduce the risk that their systems are used for illegal activity, with

specific obligations regarding fraudulent advertising, closing a key conduit exploited to scale artificial-intelligence-enabled fraud [44].

Table 2 below summarises the main uses of AI in criminal tasks.

Table 2. Artificial intelligence in criminal tasks.

Use/Tactic	Core Capability	Crimes/Examples	Why It Scales	Key Defences
Synthetic media and voice cloning	Generative adversarial networks and diffusion for video; neural text-to-speech for voices	Sextortion; senior executive impersonation; evidence tampering	High realism; low cost; reuse across channels	Content credentials and provenance standards; watermarking; trusted detection; chain-of-custody; out-of-band confirmation
Artificial-intelligence-written social engineering	Large language models for personalization; chatbot orchestration	Spear-phishing; voice and text scams; fake support chats	Personalization at machine scale	Risk-graded friction; email authentication and domain alignment; realistic drills; live call-backs
Synthetic identities and document forgery	Identity document synthesis; data correlation	Customer identity verification bypass; new-account fraud; mobile number and account swaps	Automated onboarding; high-fidelity forgeries	Strong and continuous customer verification; liveness and behavioural biometrics; intelligence sharing
Evidentiary fabrication in judicial contexts (synthetic logs, chats, emails)	Generative text with metadata/header synthesis; emulation of forensic artefacts and document provenance cues	Fabricated WhatsApp/Telegram conversation databases; forged email threads with plausible headers; synthetic server logs and transactional records	High surface plausibility; rapid, low-cost production; cross-format internal consistency that can withstand superficial scrutiny	Provenance-first acquisition and preservation; deep digital forensics (device, file system, and server-log coherence); validated methods with transparent uncertainty reporting; strict chain-of-custody and corroboration requirements
Adaptive malware and adversarial machine learning	Evasion learning; adversarial examples; data poisoning	Antivirus and sandbox evasion; polymorphic payloads	Rapid adaptation to detectors	Adversarially robust models; behaviour analytics; telemetry sharing
Botnets and automated abuse	Traffic shaping, targeting, and scheduling	Denial-of-service; large-scale scraping; credential stuffing	Real-time coordination	Bot management; rate limits; anomaly detection; proof-of-work on high-risk endpoints
Payment fraud and account takeover	Voice cloning with conversational engines	Authorised push-payment fraud; one-time password harvesting; remote-access scams	Believable prompts; urgency manipulation	Strong customer authentication; confirmation of payee; velocity controls; delayed settlement

Table 2. *Cont.*

Use/Tactic	Core Capability	Crimes/Examples	Why It Scales	Key Defences
Anti-money-laundering evasion and laundering	Normal-pattern mimicry; synthetic transaction graphs	Layering; mule networks; cross-platform washing	Low-signal, threshold-evading flows	Graph and sequence models; continuous baselining; human-in-the-loop reviews; sharing of suspicious activity reports
Organised-crime logistics and smuggling optimisation	Predictive and geospatial analytics; autonomy in drones and vehicles	Route optimisation; drone drops; surveillance evasion	Data-driven coordination; remote operations	Counter-drone systems; targeted interdiction; lawful data sharing and joint operations
Information operations and market abuse	Generative content and coordinated inauthentic behaviour	Influence operations; pump-and-dump; reputational sabotage	Platform-scale reach; rapid narrative testing	Provenance and labels; notice-and-action takedown; detection of coordination patterns; media literacy
Biometric spoofing and counter-forensics	Presentation, replay, and ultrasonic attacks; data laundering	Defeating face and voice checks; identity fraud; evidence laundering	Reusable toolkits across targets	Presentation attack detection; continuous risk scoring; multi-factor authentication with independent modalities

6. Illustrative Cases of AI-Facilitated Offending

This section surveys illustrative cases that evidence the operational heterogeneity of AI-enabled criminality and help delineate a paradigm of "assisted technological deviance." The aim is analytical rather than epidemiological: the cases are not presented as a representative prevalence sample, but as mechanism-revealing exemplars spanning synthetic-media abuse, AI-assisted cyber operations, access/infiltration via socio-technical pipelines, and the judicial/forensic risks posed by high-opacity algorithmic systems. Consistent with criminological practice in rapidly evolving domains, several cases are necessarily documented through grey literature (e.g., threat-intelligence reporting, investigative journalism, institutional advisories) because incident-level details and procedural milestones often appear first—and sometimes exclusively—outside peer-reviewed venues. Where this evidentiary base is predominantly non-academic, the case is used to characterise emerging tactics and governance vulnerabilities rather than to support quantitative claims about incidence or effect size.

To enhance pattern visibility across heterogeneous incident types, Table 3 synthesises the illustrative cases by mapping primary AI capability to offence setting, the main criminogenic implication, and the evidentiary or governance vulnerability highlighted by each case.

Table 3. Illustrative cases of AI-facilitated offending: primary capability, offence setting, and key implications.

Case	Primary AI Capability	Offence Type/Setting	Key Implication	Sources
1. University of Hong Kong deepfake ring	Synthetic image generation	Non-consensual intimate deepfakes	Consent, privacy, and institutional duty of care in high-trust environments	[116–118]
2. Maryland school audio deepfake	Voice generation	Defamation, social destabilisation	High social harm with imperfect statutory fit; evidentiary governance gaps	[119–121]
3. UK v. Hugh Nelson (synthetic CSAM)	Synthetic media generation	Creation/distribution of AI-generated CSAM	Courts treating synthetic production as materially harmful despite lack of identifiable child victim	[122–124]
4. Anthropic report on Claude Code misuse	LLM-assisted coding/ops support	Cybercrime enablement, extortion workflows	Generative AI compresses attack cycles and lowers expertise thresholds; dual-use governance limits	[125,126]
5. North Korean employment infiltration	LLM-assisted writing/translation/coding	Access via hiring pipelines	Hiring as an attack surface; insider-risk amplification through AI support	[127–129]
6. PromptLock ransomware prototype	Code generation	Proof-of-concept adaptive malware	Reduced skill barrier and faster iteration; early signal of adversarial adaptation	[130,131]
7. Ohio "Cybercheck"	AI-supported investigative analytics	Forensic AI in criminal proceedings	Opacity, validation deficits, and evidentiary admissibility risks	[132–134]
8. Facial recognition misidentifications	Computer vision	Policing and arrests	Due-process risks; need for corroboration, auditability, and oversight	[135,136]
9. 2025 synthetic-media fraud surge	Synthetic media + social engineering	Financial fraud and disinformation	Cross-platform impersonation; scale and urgency cues erode traditional safeguards	[137–139]
10. AI-enabled criminal drones	Autonomy/targeting support	Physical-domain criminal operations	Attack surface expansion from cyber to kinetic; operational interdiction challenges	[140–142]
11. AI in trafficking recruitment	ML-driven profiling/targeting	Grooming and recruitment	Industrialisation of coercive persuasion; scalable victim selection and scripting	[143–145]

1. University of Hong Kong Deepfake Ring

A law student allegedly generated more than 700 non-consensual pornographic deepfakes of fellow students and faculty. The case was initially surfaced through victim reporting on social media and prompted formal institutional responses, followed by regulatory attention. Public statements from the University indicate that internal handling included disciplinary measures and support arrangements for affected students, while broader reporting indicates that Hong Kong's privacy watchdog opened a criminal investigation, foregrounding the unlawful processing and disclosure of sensitive personal data (including intimate imagery and biometric identifiers). The incident spotlights the governance problem of consent in AI-mediated image fabrication and the duty-of-care obligations of high-trust institutions faced with low-cost synthetic-media victimisation [116–118].

2. Maryland School Audio Deepfake (Dazhon Darien)

An ex–high school athletic director used generative AI to produce a defamatory audio deepfake of the principal containing racist and antisemitic content. The recording disseminated rapidly through social platforms, precipitating threats and community unrest and producing substantial institutional disruption. The case culminated in an Alford plea and a custodial sentence, often cited as illustrative of a persistent governance mismatch: the social destructiveness and speed of synthetic-audio defamation can substantially exceed the calibration of existing criminal categories and evidentiary routines, particularly when attribution and provenance are contested early in an investigation [119–121].

3. UK v. Hugh Nelson (AI-Generated Child Sexual Abuse Material "CSAM")

Hugh Nelson received a lengthy custodial sentence for creating AI-enabled/computer-generated child sexual abuse material and distributing it. Official prosecutorial and police communications emphasise that criminal liability and legally cognizable harm may attach even where no identifiable real-world child victim can be directly mapped onto a specific synthetic output, reaffirming the protection of dignity and the preventive logic of sexual-offence frameworks in virtualized contexts. The ruling signals a stringent judicial posture toward synthetic CSAM and commercial distribution and indicates a willingness to treat tool-enabled fabrication as comparable—in legal seriousness—to traditional production modalities [122–124].

4. Anthropic Report on Claude Code Misuse

Anthropic's 27 August 2025 report describes criminal misuse of its Claude Code model in high-complexity cybercrime contexts, including malicious coding support and operational enablement for extortion workflows, with some activity attributed to North Korean-linked actors. As corporate threat-intelligence, the report does not constitute a prevalence estimate; however, it is analytically valuable because it details how generative systems can compress attacker "work cycles" by accelerating reconnaissance, automating credential-harvesting steps, and supporting adaptive decision-making. The report also describes the use of AI for psychologically tailored social engineering ("vibe hacking"), underlining that the criminogenic shift is not only technical but interactional: coercion scripts and personalization can be iterated rapidly to exploit human vulnerabilities at scale [125,126].

5. North Korean Employment Infiltration via AI

Operators linked to North Korea allegedly leveraged generative AI to draft job applications, translate materials, and develop code, securing remote roles that can provide access to corporate systems and sensitive datasets. In this pattern, hiring pipelines become an attack surface: AI reduces linguistic and cultural barriers to deception, increases the plausibility of applicant materials, and can provide "on-the-job" assistance once access is obtained, thereby amplifying insider-risk dynamics. While some operational details circulate in corporate reporting, official U.S. government communications describe the broader scheme of DPRK-linked remote IT worker fraud and associated enforcement responses, supporting the inference that employment-mediated access is an increasingly salient socio-technical vector in AI-enabled offending [127–129].

6. PromptLock: AI-Powered Ransomware Prototype

ESET researchers described "PromptLock," a proof-of-concept ransomware prototype that uses AI to generate or vary malicious code. Even if not documented as widespread deployment, its analytic relevance lies in demonstrating a plausible pathway to adversarial adaptation: code variation can erode signature-based detection and shorten attacker iteration loops, lowering expertise thresholds and accelerating learning. As an early warning signal, the case supports the broader concern that defensive controls must assume adaptive, model-assisted malware evolution rather than static toolchains [130,131].

7. Ohio's Cybercheck Tool: From Convictions to Exclusion

The Global Intelligence "Cybercheck" system was promoted as an AI-enabled investigative tool capable of geolocation inference and was reportedly used in support of prosecutions, including severe custodial outcomes. Subsequent scrutiny challenged reliability and transparency, and courts restricted or excluded Cybercheck-derived evidence in at least some proceedings, with disputes focusing on validation, disclosure, and the feasibility of meaningful adversarial testing when the analytic pipeline is opaque. This episode illustrates a central governance risk of forensic AI: where methods, error rates, and inferential steps are not independently auditable, algorithmic outputs can undermine rather than strengthen evidentiary integrity, especially when they are treated as determinative rather than as leads requiring corroboration [132–134].

8. Facial Recognition Misidentifications in U.S. Policing

Investigative reporting has documented multiple wrongful arrests linked to uncorroborated facial recognition outputs, including prolonged pretrial detention following faulty matches. The pattern raises due-process concerns where probable cause, investigative direction, or witness procedures are substantially shaped by algorithmic suggestions—particularly under conditions of "automation bias" and limited disclosure regarding tool use. The cases strengthen calls for corroboration requirements, audit trails, documentation duties, and clear thresholds for when algorithmic identification may be used at all, especially given well-described risks of demographic bias and error amplification in operational settings [135,136].

9. 2025 Surge in Synthetic-Media Fraud

Deepfake and voice-cloning scams have been reported as increasing sharply, with fraud moving beyond email into calls, video meetings, and cross-platform messaging designed to manufacture urgency and exploit trust cues. Widely reported incidents include high-value transfer fraud induced via deepfake-mediated interaction, illustrating that synthetic media can erode traditional warning signs (accent, visual inconsistency, "off-script" behaviour) by approximating familiar voices and faces under time pressure. Given that many incident narratives originate in corporate disclosure and journalism, these cases are best interpreted as evidence of tactic maturation and diffusion, not as stable estimates of incidence; nonetheless, they underscore the systemic challenge for financial governance, where verification must be resilient to impersonation at the sensory level [137–139].

10. AI-Enabled Criminal Drones in Latin America

Criminal groups have reportedly employed drones for narcotics transport, surveillance of law enforcement, and kinetic attacks, reflecting an expansion of the attack surface from cyber-enabled deception to physical-domain operations supported by semi-autonomous or remotely coordinated platforms. Open-source reporting has described explosive-equipped drone incidents and drone-supported criminal reconnaissance, indicating that lightweight autonomy can complicate interdiction and protective operations by increasing standoff distance, reducing operator exposure, and enabling rapid tactical adaptation. This pattern is best treated as an emerging operational modality rather than a quantified trend, but it is consistent with institutional analyses of how unmanned systems can be appropriated by non-state actors [140–142].

11. AI in Human Trafficking Recruitment

Trafficking networks increasingly exploit platform-mediated environments to identify and manipulate vulnerable targets. Institutional reporting indicates that machine-learning–assisted profiling can support scalable victim selection, message optimisation, and escalation scripting, effectively industrialising grooming while masking perpetrators behind algorithmically tailored personas. Even where granular incident-level data remains scarce in peer-reviewed literature, the convergence of platform affordances, targeting analytics,

and coercive persuasion techniques is a plausible mechanism for accelerating recruitment and reducing the visibility of perpetrator coordination, thereby complicating detection and victim protection efforts [143–145].

7. Discussion

AI should be understood less as a single criminal "tool" and more as a bundle of capabilities—generation, prediction, coordination, and evasion—that can be inserted across the crime script from reconnaissance to monetization and concealment. This helps explain why AI-enabled offending scales rapidly: it lowers skill thresholds, compresses attack cycles, and increases adaptability. The co-adaptive character of this landscape is visible in adversarial examples and data poisoning that degrade detection models [4,5] and in AI-authored social engineering that can match or exceed human persuasion [6,7]. Accordingly, static signatures and one-off model deployments are structurally fragile in a setting where both sides learn and update [11], consistent with law-enforcement assessments that frame AI as an accelerant for organised and transnational scams [3].

Deepfakes and voice cloning illustrate how technical maturity translates into criminogenic opportunity. Systems can synthesise identity cues from limited data, enabling impersonation, sextortion, reputational sabotage, and evidentiary manipulation [12,13,20]. Human detection of synthetic speech remains weak [21], while technical detectors often lose robustness under realistic acoustics, linguistic variation, laundering, and adversarial perturbations [29,30,39]. For criminalistic practice, the implication is that evidentiary reliability cannot be grounded in surface plausibility; it increasingly requires provenance-oriented evaluation, auditable workflows, and explicit uncertainty reporting when analytic methods are applied [56]. In parallel, financial and identity fraud shows AI as scalable deception infrastructure—tailored lures, synthetic documents, synthetic identities, and automation that overwhelms review [6,7,58,146]—while organised crime integrates these capabilities into logistics, laundering, and influence operations, exploiting blind spots and emulating "normality" to erode controls over time [3,87,95–97,113].

Legal responses are expanding but uneven. The EU's Artificial Intelligence Act and Digital Services Act provide a framework for transparency, traceability, and intermediary duties relevant to manipulated content and fraud vectors [45,46], while the UK Online Safety Act and related reforms strengthen platform responsibilities [44]. Italy's Law 23 September 2025 No. 132 explicitly links aggravated liability to AI use [114]. China's Law Against Telecom and Online Fraud (2022) [84], Singapore's Online Criminal Harms Act (2023) [85], and enforcement pathways addressing synthetic-voice robocalls under U.S. telemarketing law [83] illustrate a broader shift toward disruption powers and intermediary obligations. However, regulatory measures can create false reassurance when they are not stress-tested under adversarial adaptation, and cross-border enforcement often lags the modularity and velocity of "crime-as-a-service" ecosystems.

The criminalistic and criminological implications are best captured by integrating routine activity theory, rational choice perspectives, and techniques of neutralisation. First, across sexual exploitation [116–118,122–124], reputational harm [119–121], fraud and extortion [125,126,130,131,137–139], labour-market infiltration [127–129], and trafficking recruitment [143–145], AI functions less as an autonomous agent than as an amplifier and coordinator of human intent. Models lower skill thresholds, compress cycles, and enable fine-grained tailoring of harm while supplying templates, scaffolds, and modular services. This is most evident when generative systems provide decision support—sequencing, targeting, optimisation—so offenders can outsource reconnaissance, message craft, and iterative strategy [125,126]. The result is a migration from artisanal deviance to semi-automated workflows that are repeatable for novices and scalable for organised actors.

Second, routine activity dynamics shift. Motivated offenders are upskilled by systems that encode and disseminate tacit know-how; suitable targets become more legible through datafication and cross-platform fusion [137–139,143–145]; and capable guardianship is blunted by both scale and verisimilitude. Traditional cues for threat recognition—voice, video, institutional formatting—are degraded by synthetic media that is credible by default. Guardianship can also be weakened internally: the Cybercheck episode [132–134] and facial-recognition misidentifications [135,136] show how high-opacity tools can shape policing and adjudication in ways that are hard to audit ex ante and costly to remediate ex post, raising the risk of algorithmically assisted probable cause without adequate validation and contestability.

Third, rational choice considerations clarify why AI changes the expected utility calculus of offending. Offenders can test and iterate cheaply, probe controls, and shift tactics as detectors and policies change. This is reinforced by criminogenic affordances of AI systems: scale without proportional effort once a pipeline is operational, personalization at scale via segmentation and susceptibility targeting, and opacity in both directions as offenders exploit blind spots while institutions struggle to audit provenance and reliability [132–136]. These affordances recompose the crime script by modularizing tasks that once required durable conspiracies or specialised skills, widening participation and accelerating diffusion.

Fourth, techniques of neutralisation become easier to deploy. AI can diffuse agency and increase psychological distance, enabling denial of responsibility ("the model generated it"), denial of injury (including claims that "no real child was harmed" in synthetic CSAM contexts, notwithstanding the judicial stance in case 3 [122–124]), and normalisation through perceived ubiquity; where state-aligned actors are involved, appeals to higher loyalties may further lower internal constraints [125–129]. Victimology likewise departs from analogue baselines: synthetic sexual image fabrication [116–118] and synthetic CSAM [122–124] show that victimisation is no longer contingent on pre-existing artefacts or physical proximity, while harms span dignity violations, reputational and economic losses [119–121,137–139], coercive grooming dynamics [143–145], and collective erosion of epistemic trust [137–139].

These findings imply that effective control requires calibrated friction and procedural safeguards that anticipate adversarial adaptation. Ex ante measures include risk-tiered access controls, realistic red-teaming tied to offender scripts, and provenance infrastructures for high-risk content and transactions. Ex post measures include independent validation and error-rate disclosure for forensic AI, mandatory corroboration when algorithmic outputs inform probable cause, standardised incident and near-miss reporting, and cross-border cooperation aligned to cloud-native evidence flows. Without such measures, AI-assisted deviance risks becoming a default condition, eroding both safety and the legitimacy of criminal procedure.

8. Conclusions

AI has shifted cyber-enabled crime from artisanal deception to scalable, adaptive operations. Deepfakes and voice cloning erode identity assurance and evidentiary trust; generative tooling industrialises phishing and synthetic-identity fraud; and organised crime increasingly embeds predictive analytics to optimise logistics, target victims, and evade surveillance. In this co-adaptive threat landscape, purely reactive detection is structurally insufficient: defensive systems are repeatedly stress-tested, bypassed, and re-targeted as offenders learn detector contours and exploit governance gaps.

Accordingly, durable responses require a layered approach that couples technical measures with procedural and legal safeguards. At the technical and operational level, priority should be given to provenance-first infrastructures, continuously evaluated detec-

tion under realistic adversarial conditions, calibrated friction for high-risk transactions and identity events, and timely cross-sector intelligence sharing. At the institutional and legal level, policy should clarify platform, telecom, and payment-provider duties, strengthen admissibility standards for AI-mediated evidence through validation and transparency requirements, and introduce targeted offences and AI-specific aggravators where automation measurably scales harm.

These measures are most effective when treated as complementary rather than substitutable. Together, they can raise attacker effort and risk, reduce criminal rewards by tightening monetization choke points, protect victims' dignity and assets, and preserve evidentiary integrity while remaining consistent with fundamental rights and due process.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/sci8030054/s1, Table S1. SANRA Score for quality assessment of selected narrative review studies (n = 18) for the review.

Author Contributions: Conceptualization, P.B., G.R. and G.V.; methodology, G.N.; validation, G.N., A.S. and T.S.; formal analysis, P.B.; investigation, A.S. and G.N.; data curation, G.P. and G.V.; writing—original draft preparation, P.B. and G.P.; writing—review and editing, P.B., A.S. and T.S.; supervision, G.R. and A.S.; project administration, G.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Acknowledgments: During the preparation of this manuscript/study, the author(s) used DeepL and ChatGPT 5.1 for the purposes of enhance fluency, syntax, and grammar (text editing). The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AML	Anti-Money Laundering
C2PA	Coalition for Content Provenance and Authenticity (content credentials)
CNN(s)	Convolutional Neural Network(s)
CSAM	Child Sexual Abuse Material
DDoS	Distributed Denial of Service
GAN(s)	Generative Adversarial Network(s)
KYC	Know Your Customer
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
OTP	One-Time Passwords
PINs	Personal Identification Numbers
RNN(s)	Recurrent Neural Network(s)

References

1. Brundage, M.; Avin, S.; Clark, J.; Toner, H.; Eckersley, P.; Garfinkel, B.; Dafoe, A.; Scharre, P.; Zeitzoff, T.; Filar, B.; et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv* **2018**, arXiv:1802.07228. [CrossRef]
2. Alansary, S.A.; Ayyad, S.M.; Talaat, F.M.; El-Kafrawy, P.M.; El-Sayed, A.M.; Khafaga, D.S.; El-Sayed, H.M. Emerging AI threats in cybercrime: A review of zero-day attacks via machine, deep, and federated learning. *Knowl. Inf. Syst.* **2025**, *67*, 10951–10987. [CrossRef]
3. Europol. EU Serious and Organised Crime Threat Assessment (EU-SOCTA 2025). Available online: https://www.europol.europa.eu/publications-events/main-reports/socta-report (accessed on 18 October 2025).
4. Patil, S.; Varadarajan, V.; Walimbe, D.; Patil, B. Improving the robustness of AI-based malware detection using adversarial machine learning. *Algorithms* **2021**, *14*, 297. [CrossRef]
5. Chen, S.; Xue, M.; Fan, L.; Hao, S.; Xu, L.; Li, B. Automated poisoning attacks and defenses in malware detection systems: An adversarial machine learning approach. *arXiv* **2017**, arXiv:1706.04146. [CrossRef]
6. Schmitt, M.; Flechais, I. Digital deception: Generative artificial intelligence in social engineering and phishing. *Artif. Intell. Rev.* **2024**, *57*, 324. [CrossRef]
7. Heiding, F.; Lermen, S.; Kao, A.; Miettinen, M.; Sasse, A.; Asokan, N. Evaluating large language models' capability to launch fully automated spear phishing campaigns: Validated on human subjects. *arXiv* **2024**, arXiv:2412.00586. [CrossRef]
8. Jabir, R.; Le, J.; Nguyen, C. Phishing attacks in the age of generative artificial intelligence: A systematic review of human factors. *AI* **2025**, *6*, 174. [CrossRef]
9. Goyal, P.; Hossain, K.S.M.T.; Deb, A.; Ferraro, A.; Wellman, M.P. Discovering signals from web sources to predict cyber attacks. *arXiv* **2018**, arXiv:1806.03342. [CrossRef]
10. Danish, M. Enhancing cyber security through predictive analytics: Real-time threat detection and response. *arXiv* **2024**, arXiv:2407.10864. [CrossRef]
11. Alavizadeh, H.; Jang-Jaccard, J.; Alpcan, T.; Doss, R.; Camtepe, S. A Markov game model for AI-based cybersecurity attack mitigation. *arXiv* **2021**, arXiv:2107.09258.
12. Umbach, R.; Henry, N.; Beard, G.; Johnson, K.M.; Ringrow, J.; Scott, M.; Pina, A.; Sanches, D.; Gámez-Guadix, M.; Koymen, B.; et al. Non-consensual synthetic intimate imagery: Prevalence, attitudes, and knowledge in 10 countries. *arXiv* **2024**, arXiv:2402.01721. [CrossRef]
13. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. and beyond: A survey of face manipulation and fake detection. *Inf. Fusion* **2020**, *64*, 131–148. [CrossRef]
14. Sandoval, M.; De Almeida Vau, M.; Solaas, J.; Weinborn, C.; Verkruysse, W. Threat of deepfakes to the criminal justice system: A systematic review. *Crime Sci.* **2024**, *13*, 41. [CrossRef]
15. Norta, A.; Makrygiannis, S. Designing AI-equipped social decentralized autonomous organizations for tackling sextortion cases (Version 0.7). *arXiv* **2023**, arXiv:2312.14090.
16. Malik, S.; Surbhi, A.; Roy, D. Blurring boundaries between truth and illusion: Analysis of human rights and regulatory concerns arising from abuse of deepfake technology. *AIP Conf. Proc.* **2024**, *3220*, 050016. [CrossRef]
17. Chapagain, D.; Kshetri, N.; Aryal, B. Deepfake Disasters: A Comprehensive Review of Technology, Ethical Concerns, Countermeasures, and Societal Implications. In Proceedings of the 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Windhoek, Namibia, 23–25 July 2024; pp. 1–9.
18. Nasution, A.V.A.; Suteki, N.; Lumbanraja, A.D. Addressing Deepfake Pornography and the Right to be Forgotten in Indonesia: Legal Challenges in the Era of AI-Driven Sexual Abuse. *Int. J. Semiot. Law* **2025**, *38*, 2489–2517. [CrossRef]
19. Wang, Y.; Skerry-Ryan, R.; Stanton, D.; Wu, Y.; Weiss, R.J.; Jaitly, N.; Yang, Z.; Xiao, Y.; Chen, Z.; Bengio, S.; et al. Tacotron: Towards End-to-End Speech Synthesis. *arXiv* **2017**, arXiv:1703.10135.
20. Prenger, R.; Valle, R.; Catanzaro, B. WaveGlow: A flow-based generative network for speech synthesis. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3617–3621.
21. Mai, K.T.; Bray, S.; Davies, T.; Griffin, L.D. Warning: Humans cannot reliably detect speech deepfakes. *PLoS ONE* **2023**, *18*, e0285333. [CrossRef]
22. Wang, K.; Chen, M.; Lu, L.; Li, J.; Zhang, Y.; Zhang, K.; Li, Q.; Wang, X. From one stolen utterance: Assessing the risks of voice cloning in the AIGC era. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 12–15 May 2025; pp. 4663–4681.
23. Allagi, S.; Preethi, P.; Rodriguez-Baca, L.S.; De La Vega, C.F.C.P. A Voice Cloning Detection: Distinguishing Real from Mimicked Voices Using SVM. In Proceedings of the 2025 International Conference on Computing Technologies (ICOCT), Bengaluru, India, 13–14 June 2025; pp. 1–4.
24. Lakshmanan, V.; Ferri, D.; Agarwal, R.; Beling, P.A. Evaluating the Importance of Demographic and Technical Factors in Creating Authentic-Sounding AI-Generated Human Voice Clones. In Proceedings of the Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2 May 2025; pp. 1–6.

25. Milewski, K.; Zaporowski, S.; Czyżewski, A. Comparison of the ability of neural network model and humans to detect a cloned voice. *Electronics* **2023**, *12*, 4458. [CrossRef]
26. Kassis, A.; Hengartner, U. Breaking Security-Critical voice authentication. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 22–25 May 2023; pp. 951–968.
27. Kheria, I.; Karani, R. CloneAI: A Deep Learning-Based Approach for Cloned Voice Detection. In *Data Science and Applications*; Nanda, S.J., Yadav, R.P., Gandomi, A.H., Saraswat, M., Eds.; Lecture Notes in Networks and Systems; Springer: Singapore, 2024; Volume 820, pp. 267–282.
28. Ali, H.; Subramani, S.; Sudhir, S.; Varahamurthy, R.; Malik, H. Is Audio Spoof Detection Robust to Laundering Attacks? In Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security (IH & MMSec '24), Baiona, Spain, 24–26 June 2024; pp. 283–288.
29. Gao, C.; Postiglione, M.; Gortner, I.; Kraus, S.; Subrahmanian, V.S. Perturbed Public Voices (P^2V): A dataset for robust audio deepfake detection. *arXiv* **2025**, arXiv:2508.10949.
30. Nguyen, B.; Shi, S.; Ofman, R.; Le, T. What you read isn't what you hear: Linguistic sensitivity in deepfake speech detection. *arXiv* **2025**, arXiv:2505.17513.
31. Li, X.; Li, N.; Zhong, J.; Liu, Q.; Lee, K.A. Investigating robustness of adversarial samples detection for automatic speaker verification. In Proceedings of the Interspeech 2020, Shanghai, China, 25–29 October 2020; pp. 1540–1544.
32. Nandal, A.; Dua, M. A hybrid approach to secure automatic speaker verification: Integrating clone detection and speaker identification. *Int. J. Speech Technol.* **2025**, *28*, 411–429. [CrossRef]
33. Kabir, H.M.D.; Khosravi, A.; Hosen, M.A.; Nahavandi, S. Partial adversarial training for prediction interval. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–6.
34. Go, C.; Park, N.I.; Jeon, O.; Chun, C. A Pre-Training framework based on Multi-Order acoustic simulation for replay voice spoofing detection. *Sensors* **2023**, *23*, 7280. [CrossRef]
35. Zuo, C.; Jia, Z.; Li, W. AdVTTS: Adversarial Text-to-Speech Synthesis attack on speaker identification systems. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 4840–4844.
36. Wu, X.; Ma, S.; Shen, C.; Wang, X.; Zhang, C.; Zhang, K. KENKU: Towards Efficient and Stealthy Black-box Adversarial Attacks against ASR Systems. In Proceedings of the 32nd USENIX Security Symposium (USENIX Security 2023), Anaheim, CA, USA, 9–11 August 2023; pp. 247–264.
37. Abdelhakim, M.Y.; Ahmed, D.M.K. Adversarial attacks for speaker recognition system with FGSM-CNN and FGSM-DNN. In Proceedings of the 2024 International Conference on Telecommunications and Intelligent Systems (ICTIS), Djelfa, Algeria, 14–15 December 2024; pp. 1–6.
38. Fei, Q.; Hou, W.; Hai, X.; Liu, X. VocalCrypt: Novel active defense against deepfake voice based on masking effect. *arXiv* **2025**, arXiv:2502.10329. [CrossRef]
39. Zhang, B.; Cui, H.; Nguyen, V.; Whitty, M. Audio deepfake detection: What has been achieved and what lies ahead. *Sensors* **2025**, *25*, 1989. [CrossRef] [PubMed]
40. Verdoliva, L. Media Forensics and DeepFakes: An Overview. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 910–932. [CrossRef]
41. Romero-Moreno, F. Deepfake Detection in Generative AI: A Legal Framework Proposal to Protect Human Rights. *Comput. Law Secur. Rev.* **2025**, *58*, 106162. [CrossRef]
42. AbdulQudus, A.B.; Amodu, O.A.; Bukar, U.A.; Olanrewaju, R.F.; Yusoff, M.Z.B.; Alabi, A.B. Contemporary and Comprehensive Bibliometric Exposition on Deepfake Research and Trends. *Comput. Mater. Contin.* **2025**, *84*, 153–236. [CrossRef]
43. UK Government. Government Crackdown on Explicit Deepfakes. News Release. 7 January 2025. Available online: https://www.gov.uk/government/news/government-crackdown-on-explicit-deepfakes (accessed on 18 October 2025).
44. United Kingdom. Online Safety Act 2023. Available online: https://www.legislation.gov.uk/ukpga/2023/50 (accessed on 18 October 2025).
45. European Union. Regulation (EU) 2024/1689 of 13 June 2024 (Artificial Intelligence Act). Available online: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (accessed on 18 October 2025).
46. European Union. Regulation (EU) 2022/2065 of 19 October 2022 (Digital Services Act). Available online: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng (accessed on 18 October 2025).
47. People's Republic of China. Administrative Provisions on Deep Synthesis in Internet-Based Information Services (Effective 10 January 2023). Available online: https://cyrilla.org/entity/gercxuzg62o?file=1728288988021wmhrfp2o9t.pdf (accessed on 18 October 2025).
48. eSafety Commissioner (Australia). Learn About the Online Safety Act. Available online: https://www.esafety.gov.au/newsroom/whats-on/online-safety-act (accessed on 18 October 2025).

49. Brazil. Tribunal Superior Eleitoral. Resolução No. 23.732, de 27 de Fevereiro de 2024 (Propaganda Eleitoral—IA e Deepfake). Available online: https://www.tse.jus.br/legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024 (accessed on 18 October 2025).
50. National Conference of State Legislatures (USA). Artificial Intelligence (AI) in Elections and Campaigns. Available online: https://www.ncsl.org/elections-and-campaigns/artificial-intelligence-ai-in-elections-and-campaigns (accessed on 18 October 2025).
51. Tennessee. Public Chapter No. 588 (Ensuring Likeness, Voice, and Image Security—ELVIS Act, 2024). Available online: https://publications.tnsosfiles.com/acts/113/pub/pc0588.pdf (accessed on 18 October 2025).
52. Republic of Korea. Act on Special Cases Concerning the Punishment of Sexual Crimes (English Version, Consolidated). Available online: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=40947&lang=ENG (accessed on 18 October 2025).
53. Qureshi, S.M.; Saeed, A.; Almotiri, S.H.; Masood, A.; Ghabban, F.; Alzahrani, A.J. A survey of digital forensic methods for multimodal deepfake identification on social media. *PeerJ Comput. Sci.* **2024**, *10*, e2037. [CrossRef]
54. Lin, C.-Y.; Lee, J.-C.; Wang, S.-J.; Chang, C.-C. Video detection method based on temporal and spatial foundations for accurate verification of authenticity. *Electronics* **2024**, *13*, 2132. [CrossRef]
55. Amerini, I.; Barni, M.; Battiato, S.; Bestagini, P.; Cozzolino, D.; Delp, E.J.; Farinella, G.M.; Fioretti, F.; Galdi, C.; Giudice, O.; et al. Deepfake media forensics: Status and future challenges. *J. Imaging* **2025**, *11*, 73. [CrossRef] [PubMed]
56. Coalition for Content Provenance and Authenticity (C2PA). C2PA Technical Specification. Available online: https://c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html (accessed on 18 October 2025).
57. Tietoevry Banking. Banking's New Insight-Report Reveals an Increase in Digital Payment Fraud in Europe. Available online: https://www.tietoevry.com/en/newsroom/all-news-and-releases/press-releases/2025/04/tietoevry-bankings-new-insight-report-reveals-an-increase-in-digital-payment-fraud-in-europe (accessed on 18 October 2025).
58. Experian UK & Ireland. New Figures from Experian Reveal the Rise of 'Synthetic Fraud' in the UK. Available online: https://www.experianplc.com/newsroom/press-releases/2025/-synthetic-fraud--reaches-record-levels? (accessed on 18 October 2025).
59. Payments Association. Synthetic Identity Fraud in the Age of AI: Why Payments Needs a Global, Risk-Based Response. Available online: https://thepaymentsassociation.org/article/synthetic-identity-fraud-in-the-age-of-ai-why-payments-needs-a-global-risk-based-response/? (accessed on 18 October 2025).
60. Finextra. AI-Created Digital Documents and Deep Fakes Pose Biggest Threat to Financial Services. Available online: https://www.finextra.com/newsarticle/45077/ai-created-digital-documents-and-deep-fakes-pose-biggest-threat-to-financial-services (accessed on 18 October 2025).
61. Maharana, N.; Kuppili, S.K.; Ganesh, B.U.B.; Patro, K.K. From Defense to Deception. In *Generative AI for Business Analytics and Strategic Decision Making in Service Industry*; Advances in Business Strategy and Competitive Advantage; IGI Global: Hershey, PA, USA, 2025; pp. 317–340.
62. Ismail, M.M.; Haq, M.A. Enhancing Enterprise Financial Fraud Detection Using Machine Learning. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 14854–14861. [CrossRef]
63. Desrousseaux, R.; Bernard, G.; Mariage, J. Profiling Money Laundering with Neural Networks: A Case Study on Environmental Crime Detection. In Proceedings of the 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), Washington, DC, USA, 1–3 November 2021; pp. 364–369.
64. Ahmadi, S.; Mazjini, M. Generative AI in Fraud Prevention. In *Generative AI Insights for Financial Decision Making*; Advances in Computational Intelligence and Robotics; IGI Global: Hershey, PA, USA, 2025; pp. 129–158.
65. Lyeonov, S.; Draskovic, V.; Kubaščikova, Z.; Fenyves, V. Artificial Intelligence and Machine Learning in Combating Illegal Financial Operations: Bibliometric Analysis. *Hum. Technol.* **2024**, *20*, 325–360. [CrossRef]
66. Amdouni, A. Artificial Intelligence and Financial Fraud Detection. In *Artificial Intelligence for Financial Risk Management and Analysis*; Advances in Computational Intelligence and Robotics; IGI Global: Hershey, PA, USA, 2025; pp. 271–280.
67. Prakash, V.; Deokar, R. Harnessing AI for Fraud Detection and Prevention in Finance and Banking. In *Real-World Applications of AI Innovation*; Mallik, S., Mathivanan, S.K., Sangeetha, S.K.B., Soufiene, B.O., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 389–406.
68. Balaji, K. Artificial Intelligence for Enhanced Anti-Money Laundering and Asset Recovery: A New Frontier in Financial Crime Prevention. In Proceedings of the 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, 28–30 August 2024; pp. 1010–1016.
69. Syed, F.A.; Fang, K.; Kiani, A.K.; Khan, M.A.; Khan, M.A. Novel Intelligent Supervised Neuro-Structures for Nonlinear Financial Crime Differential Systems. *Mod. Phys. Lett. B* **2024**, *39*, 2450399. [CrossRef]
70. Ismaeil, M.K.A. Harnessing AI for Next-Generation Financial Fraud Detection: A Data-Driven Revolution. *J. Ecohumanism* **2024**, *3*, 811–821. [CrossRef]
71. Wang, Z.; Shen, Q.; Bi, S.; Zhang, Y.; Li, X. AI Empowers Data Mining Models for Financial Fraud Detection and Prevention Systems. *Procedia Comput. Sci.* **2024**, *243*, 891–899. [CrossRef]

72. Shafik, W. The Role of Generative Artificial Intelligence in E-Commerce Fraud Detection and Prevention. In *Strategies for E-Commerce Data Security: Cloud, Blockchain, AI, and Machine Learning*; Advances in Web Technologies and Engineering; IGI Global: Hershey, PA, USA, 2024; pp. 430–469.
73. Mhammad, A.F.; Agarwal, R.; Columbo, T.; Alsmadi, I. Generative & Responsible AI—LLMs Use in Differential Governance. In Proceedings of the 2023 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 13–15 December 2023; pp. 291–295.
74. Borkar, B.S.; Patil, D.R.; Markad, A.V.; Patil, P.B. Real or Fake Identity Deception of Social Media Accounts Using Recurrent Neural Network. In Proceedings of the 2022 International Conference (ICFIRTP 2022), Roorkee, Uttarakhand, India, 23–24 November 2022.
75. Patil, D.R.; Pattewar, T.M.; Punjabi, V.D.; Shinde, G.R. Detecting Fake Social Media Profiles Using the Majority Voting Approach. *EAI Endorsed Trans. Scalable Inf. Syst.* **2024**, *11*, 4264. [CrossRef]
76. Buccafurri, F.; Lax, G.; Migdal, D.; Nocera, A.; Ursino, D. Contrasting False Identities in Social Networks by Trust Chains and Biometric Reinforcement. In Proceedings of the 2017 International Conference on Cyberworlds (CW), Chester, UK, 20–22 September 2017; pp. 17–24.
77. Bullock, H.; Edwards, M. Temporal Constraints in Online Dating Fraud Classification. In Proceedings of the 9th International Conference on Information Systems Security and Privacy (ICISSP 2023), Lisbon, Portugal, 22–24 February 2023; pp. 535–542.
78. Bahri, L.; Carminati, B.; Ferrari, E. Community-Based Identity Validation on Online Social Networks. In Proceedings of the 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCS Workshops), Madrid, Spain, 30 June–3 July 2014; pp. 21–30.
79. Vallarino, D. Detecting Financial Fraud with Hybrid Deep Learning: A Mix-of-Experts Approach to Sequential and Anomalous Patterns. *arXiv* **2025**, arXiv:2504.03750.
80. He, S.; Lei, Y.; Zhang, Z.; Li, S.; Zhang, Y.; Wang, C. Identity Deepfake Threats to Biometric Authentication Systems: Public and Expert Perspectives. *arXiv* **2025**, arXiv:2506.06825. [CrossRef]
81. European Union. Directive (EU) 2015/2366 of 25 November 2015 on Payment Services in the Internal Market. Available online: https://eur-lex.europa.eu/eli/dir/2015/2366/oj/eng (accessed on 18 October 2025).
82. United Kingdom. Economic Crime and Corporate Transparency Act 2023. Available online: https://www.legislation.gov.uk/ukpga/2023/56 (accessed on 18 October 2025).
83. United States, Federal Communications Commission. Declaratory Ruling FCC 24-17. 8 February 2024. Available online: https://docs.fcc.gov/public/attachments/FCC-24-17A1.pdf (accessed on 18 October 2025).
84. People's Republic of China. Law Against Telecom and Online Fraud (Effective 1 December 2022). Available online: https://en.moj.gov.cn/2023-12/15/c_948363.htm (accessed on 18 October 2025).
85. Singapore. Online Criminal Harms Act 2023 (No. 24 of 2023). Available online: https://sso.agc.gov.sg/Act/OCHA2023 (accessed on 18 October 2025).
86. Australia. Anti-Money Laundering and Counter-Terrorism Financing Act 2006. Available online: https://www.legislation.gov.au/Details/C2020C00362 (accessed on 18 October 2025).
87. Racoveanu, C.C. Artificial intelligence—A double-edged sword: Organized crime's AI vs law enforcement's AI. *Proc. Int. Conf. Bus. Excell. PICBE* **2024**, *18*, 445–454. [CrossRef]
88. Yang, C. CrimeGAT: Leveraging graph attention networks for enhanced predictive policing in criminal networks. *arXiv* **2023**, arXiv:2311.18641. [CrossRef]
89. Orogun, O.; Ogungbe, L.; Adegboye, N.; Ogunwobi, O.O. Strategies for combating synthetic identity fraud: The role of machine learning and behavioral analysis in enhancing financial ecosystem security. *Int. J. Res. Eng. Sci.* **2024**, *12*, 280–292.
90. Kaushik, P.; Garg, V.; Priya, A.; Rani, R. Financial fraud and manipulation. In *Deepfakes and Their Impact on Business*; Advances in Business Information Systems and Analytics; IGI Global Scientific Publishing: Hershey, PA, USA, 2024; pp. 173–196.
91. Ali, M.; Fernando, Z.J.; Huda, C.; Noor, N.M. Deepfakes and victimology: Exploring the impact of digital manipulation on victims. *Substant. Justice Int. J. Law* **2025**, *8*, 306. [CrossRef]
92. Beemamol, M. Unmasking the Threat: A Viewpoint on AI-Based Deepfake Financial Crimes. In *Advancements in Cyber Crime Investigations and Modern Data Analytics*; Shandilya, S.K., Sujay, D., Gupta, V.B., Eds.; CRC Press: Boca Raton, FL, USA, 2024; pp. 123–142.
93. Ring, T. Europol: The AI hacker threat to biometrics. *Biometr. Technol. Today* **2021**, *2021*, 9–11. [CrossRef]
94. Kaur, U.; Siddhey, P.K. Deepfake prospects, mitigating factors, and deceptions. In *Deepfakes and Their Impact on Business*; Advances in Business Information Systems and Analytics; IGI Global Scientific Publishing: Hershey, PA, USA, 2024; pp. 197–220.
95. Zhang, L.; Liang, Y.; Zhou, C.; Zhang, Y.; Li, X. Smuggling crime clue mining based on logistics big data. *Front. Artif. Intell. Appl.* **2024**, *364*, 132–140.
96. ENACT Africa. AI and Organised Crime in Africa. ENACT Africa Policy Brief, 2024. Available online: https://enactafrica.org/enact-observer/ai-and-organised-crime-in-africa (accessed on 18 October 2025).

97. Institute for Security Studies (ISS Africa). Risks and Rewards of AI for Organised Crime in Africa. Policy Report, 2024. Available online: https://issafrica.org/iss-today/risks-and-rewards-of-ai-for-organised-crime-in-africa (accessed on 18 October 2025).
98. Mungai, R. Synthetic Identity Fraud: A Critical Primary National Security Priority. SSRN. 23 March 2024. Available online: https://ssrn.com/abstract=4770398 (accessed on 18 October 2025).
99. Garcia-Bedoya, O.; Granados, O.; Burgos, J.C. AI against money laundering networks: The Colombian case. *J. Money Laund. Control* **2020**, *24*, 49–62. [CrossRef]
100. Al-Ababneh, H.A.; Nuralieva, C.; Usmanalieva, G.; Zholdoshova, A. The use of artificial intelligence to detect suspicious transactions in the anti-money laundering system. *Theor. Pract. Res. Econ. Fields* **2024**, *15*, 1039. [CrossRef]
101. Bozorgi, M. Exploring the role of fintech in terrorism financing: Legal frameworks and solutions. In *Sustainability and Financial Services in the Digital Age*; Springer Proceedings in Business and Economics; Springer: Cham, Switzerland, 2024; pp. 9–20.
102. Harris, H. Artificial intelligence and policing of financial crime: A legal analysis of the state of the field. In *Financial Technology and the Law*; Law, Governance and Technology Series; Springer: Cham, Switzerland, 2022; pp. 281–299.
103. Hataley, T. Trade-based money laundering: Organized crime, learning and international trade. *J. Money Laund. Control* **2020**, *23*, 651–661. [CrossRef]
104. Pavlidis, G. AI Capone, or the criminal masterminds of the future. In *Science Fiction as Legal Imaginary*; Routledge: New York, NY, USA, 2024; pp. 185–199.
105. Johnsen, J.W.; Franke, K. Identifying central individuals in organised criminal groups and underground marketplaces. In *Computational Science—ICCS 2018*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; pp. 379–386.
106. Nwanga, M.E.; Okafor, K.C.; Achumba, I.E.; Chukwudebe, G.A.; Nwogbaga, N.C. Predictive forensic based—Characterization of hidden elements in criminal networks using Baum–Welch optimization technique. In *Illumination of Artificial Intelligence in Cybersecurity and Forensics*; Lecture Notes on Data Engineering and Communications Technologies; Springer: Cham, Switzerland, 2022; pp. 231–254.
107. Raja, M.R.; Hosen, M.A.; Kabir, M.F.; Akter, S.; Uddin, M.S. Detecting and preventing money laundering using deep learning and graph analysis. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 6. [CrossRef]
108. Gandhi, H.; Tandon, K.; Gite, S.; Thakur, N. Navigating the complexity of money laundering: Anti–money laundering advancements with AI/ML insights. *Int. J. Smart Sens. Intell. Syst.* **2024**, *17*, 24. [CrossRef]
109. Chen, Z.; Soliman, W.M.; Nazir, A.; Iqbal, M.; Abdelgawad, A.E. Variational autoencoders and Wasserstein generative adversarial networks for improving the anti-money laundering process. *IEEE Access* **2021**, *9*, 83762–83785. [CrossRef]
110. Larik, A.S.; Haider, S. Clustering-based anomalous transaction reporting. *Procedia Comput. Sci.* **2011**, *3*, 606–610. [CrossRef]
111. Han, J.; Huang, Y.; Liu, S.; Van Dijk, D. Artificial intelligence for anti-money laundering: A review and extension. *Digit. Financ.* **2020**, *2*, 211–239. [CrossRef]
112. Chitimira, H.; Torerai, E.; Jana, L. Leveraging Artificial Intelligence to Combat Money Laundering and Related Crimes in the South African Banking Sector. *Potchefstroom Electron. Law J.* **2024**, *27*, 1–30. [CrossRef]
113. Zhu, M.; Gong, Y.; Xiang, Y.; Li, X. Utilizing GANs for fraud detection: Model training with synthetic transaction data. In Proceedings of the International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024), Guangzhou, China, 1–3 March 2024; SPIE: Bellingham, WA, USA, 2024; Volume 13180, pp. 887–894.
114. Italy. Law 23 September 2025 No. 132. Disposizioni e Deleghe al Governo in Materia di Intelligenza Artificiale. Gazzetta Ufficiale, Serie Generale No. 223. 25 September 2025. Available online: https://www.gazzettaufficiale.it/eli/id/2025/09/25/25G00143/sg (accessed on 18 October 2025).
115. Singapore Police Force. Introduction to the Online Criminal Harms Act. Available online: https://www.police.gov.sg/KnowledgeHub/Legislation/Online-Harms-Act-Overview/Introduction-to-OCHA (accessed on 18 October 2025).
116. People. Law Student Allegedly Used AI to Create Porn of Fellow Students—Then Tried to Apologize. Available online: https://people.com/law-student-allegedly-used-ai-create-porn-fellow-students-11773557 (accessed on 18 October 2025).
117. South China Morning Post. University of Hong Kong Warns Student over AI-Generated Porn Pics of Classmates. Available online: https://www.scmp.com/news/hong-kong/education/article/3318010/university-hong-kong-warns-student-over-ai-generated-porn-pics-classmates (accessed on 18 October 2025).
118. Statement by HKU on Individual Student Allegedly Using AI Tools to Create Indecent Images. Available online: https://hku.hk/press/news_detail_28498.html (accessed on 5 January 2026).
119. AP News. Former School Athletic Director Gets 4 Months in Jail in Racist AI Deepfake Case. Available online: https://apnews.com/article/487ea673b0449077cb23e7970546cb9f (accessed on 18 October 2025).
120. AP News. Athletic Director Used AI to Frame Principal with Racist Remarks in Fake Audio Clip, Police Say. Available online: https://apnews.com/article/ai-artificial-intelligence-principal-audio-maryland-baltimore-county-pikesville-853ed171369bcbb888eb54f55195cb9c (accessed on 18 October 2025).
121. Athletic Director Charged in Pikesville High School AI Case. Available online: https://www.baltimorecountymd.gov/departments/police/news/athletic-director-charged-pikesville-high-school-ai-case (accessed on 5 January 2026).

122. The Guardian. Man Who Used AI to Create Child Abuse Images Jailed for 18 Years. Available online: https://www.theguardian.com/uk-news/2024/oct/28/man-who-used-ai-to-create-child-abuse-images-jailed-for-18-years (accessed on 18 October 2025).
123. Sky News. Paedophile Hugh Nelson Who Made AI Child Abuse Images from Real Pictures Sent to Him Jailed for 18 Years in 'Deeply Horrifying' Landmark Case. Available online: https://news.sky.com/story/paedophile-hugh-nelson-who-made-ai-child-abuse-images-from-real-pictures-sent-to-him-jailed-for-18-years-in-deeply-horrifying-landmark-case-13220848 (accessed on 18 October 2025).
124. Man Who Used AI Technology to Create Child Sexual Abuse Images Jailed. Available online: https://www.cps.gov.uk/cps/news/man-who-used-ai-technology-create-child-sexual-abuse-images-jailed (accessed on 5 January 2026).
125. Anthropic. Detecting and Countering Misuse of AI: August 2025. Available online: https://www.anthropic.com/news/detecting-countering-misuse-aug-2025 (accessed on 18 October 2025).
126. MRW.it. Anthropic Raises the Alarm About 'Vibe Hacking' with AI. Available online: https://www.mrw.it/news/anthropic-lancia-lallarme-contro-il-vibe-hacking-con-lai/ (accessed on 18 October 2025).
127. Cointelegraph. Criminals Are 'Vibe Hacking' with AI at Unprecedented Levels. Available online: https://cointelegraph.com/news/cybercriminals-vibe-hacking-ai-ransoms-says-anthropic (accessed on 18 October 2025).
128. Microsoft Security Blog. Jasper Sleet: North Korean Remote IT Workers' Evolving Tactics to Infiltrate Organizations. Available online: https://www.microsoft.com/en-us/security/blog/2025/06/30/jasper-sleet-north-korean-remote-it-workers-evolving-tactics-to-infiltrate-organizations/ (accessed on 18 October 2025).
129. Justice Department Disrupts North Korean Remote IT Worker Fraud Schemes Through Charges and Arrest of Nashville Facilitator. Available online: https://www.justice.gov/archives/opa/pr/justice-department-disrupts-north-korean-remote-it-worker-fraud-schemes-through-charges-and (accessed on 5 January 2026).
130. ESET. ESET Discovers PromptLock, the First AI-Powered Ransomware. Available online: https://www.eset.com/us/about/newsroom/research/eset-discovers-promptlock-the-first-ai-powered-ransomware/ (accessed on 18 October 2025).
131. CyberScoop. NYU Team Behind AI-Powered Malware Dubbed 'PromptLock'. Available online: https://cyberscoop.com/ai-ransomware-promptlock-nyu-behind-code-discovered-by-security-researchers/ (accessed on 18 October 2025).
132. WIRED. This AI Tool Helped Convict People of Murder. Then Someone Took a Closer Look. Available online: https://www.wired.com/story/cybercheck-crime-reports-prosecutions/ (accessed on 18 October 2025).
133. Business Insider. Prosecutors Used an AI Tool to Send a Man to Prison for Life. Now the Person Who Created It Is Under Investigation. Available online: https://www.businessinsider.com/ai-crime-tool-cybercheck-founder-adam-mosher-investigation-2024-8 (accessed on 18 October 2025).
134. State v. Carr, 2024-Ohio-4471. Available online: https://www.supremecourt.ohio.gov/rod/docs/pdf/9/2024/2024-Ohio-4471.pdf (accessed on 5 January 2026).
135. The Washington Post. Arrested by AI: Police Ignore Standards After Facial Recognition Matches. Available online: https://www.washingtonpost.com/business/interactive/2025/police-artificial-intelligence-facial-recognition/ (accessed on 18 October 2025).
136. Opening Statement of Ranking Member Jamie Raskin Subcommittee on Crime and Federal Government Surveillance Hearing Titled "Artificial Intelligence and Criminal Exploitation: A New Era of Risk". Available online: https://www.congress.gov/119/meeting/house/118467/documents/HHRG-119-JU08-20250716-SD007-U7.pdf (accessed on 5 January 2026).
137. Identity Theft Resource Center. 2025 Trends in Identity Report. Available online: https://www.idtheftcenter.org/post/2025-trends-in-identity-report-impersonation-scams-rise/ (accessed on 18 October 2025).
138. Reuters. Rubio Impersonator Used AI, Signal to Contact Foreign Officials. Available online: https://www.reuters.com/world/us/rubio-impersonator-used-ai-calls-foreign-ministers-cable-shows-2025-07-08/ (accessed on 18 October 2025).
139. Senior U.S. Officials Continue to be Impersonated in Malicious Messaging Campaign. Available online: https://www.ic3.gov/PSA/2025/PSA251219 (accessed on 5 January 2026).
140. Global Initiative Against Transnational Organized Crime. Crime by Drone: A New Paradigm for Organized Crime. Available online: https://globalinitiative.net/analysis/crime-by-drone-a-new-paradigm-for-organized-crime/ (accessed on 18 October 2025).
141. Small Wars Journal. Mexican Cartel Tactical Note #38: Armed Drone Targets the Baja California Public Safety Secretary. Available online: https://archive.smallwarsjournal.com/jrnl/art/mexican-cartel-tactical-note-38-armed-drone-targets-baja-california-public-safety (accessed on 18 October 2025).
142. Unmanned Futures: The Impact of Robotics and Unmanned Systems on Law Enforcement. Available online: https://www.europol.europa.eu/publication-events/main-reports/unmanned-futures (accessed on 5 January 2026).
143. Europol. Over 30 Potential Victims Identified in Action Against Human Trafficking Enabled Online. Available online: https://www.europol.europa.eu/media-press/newsroom/news/over-30-potential-victims-identified-in-action-against-human-trafficking-enabled-online (accessed on 18 October 2025).

144. Capacity4dev (European Commission). Artificial Intelligence and Organised Crime (EL PAcCTO 2.0). Available online: https://capacity4dev.europa.eu/library/artificial-intelligence-and-organised-crime-english_en (accessed on 18 October 2025).
145. New Frontiers: The Use of Generative Artificial Intelligence to Facilitate Trafficking in Persons. Available online: https://www.osce.org/sites/default/files/f/documents/7/d/579715.pdf (accessed on 5 January 2026).
146. Mahdi, M.A.; Arshed, M.A.; Muneer, A. One Model for Many Fakes: Detecting GAN and Diffusion-Generated Forgeries in Faces, Invoices, and Medical Heterogeneous Data. *Mathematics* **2025**, *13*, 3093. [CrossRef]