Global Journal of Politics and Law Research, 14(2),68-79, 2026
ISSN: ISSN 2053-6321(Print),
ISSN: ISSN 2053-6593(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development–UK

Artificial Intelligence-Driven Cybercrime: Emerging Threats and Implications for Digital Security Governance

Kelvin Bribena
Faculty of Law, Niger Delta University, Wilberforce Island, Bayelsa State

doi: https://doi.org/10.37745/gjplr.2013/vol14n26879 Published May 02, 2026

Citation: Bribena K. (2026) Artificial Intelligence-Driven Cybercrime: Emerging Threats and Implications for Digital Security Governance, *Global Journal of Politics and Law Research*, 14(2),68-79

Abstract: *This article examines artificial intelligence-driven cybercrime as an emerging threat to digital security governance, with particular focus on AI-enabled phishing, deepfakes and automated cyberattacks. The study argues that artificial intelligence has transformed cybercrime by increasing the speed, scale, realism and sophistication of digital offences. AI tools can generate convincing phishing messages, clone voices, create deepfake videos, automate vulnerability scanning and support large-scale social engineering attacks. These developments create serious challenges for cybercrime law, digital evidence, institutional cybersecurity and public trust. The article adopts a doctrinal and analytical approach by reviewing legal frameworks, cybercrime literature and emerging AI governance responses. It finds that existing cybercrime laws provide an important foundation for addressing fraud, identity theft, unauthorised access and data misuse, but they may not fully respond to the distinctive risks created by synthetic identities, AI-generated deception and automated attacks. The study further shows that deepfakes complicate digital evidence by weakening the reliability of audio, image and video materials, while AI-enabled phishing increases the vulnerability of individuals and organisations to fraud. The article concludes with recommendations that AI-driven cybercrime should be addressed through a broader digital security governance framework. This requires stronger cybercrime laws, improved digital forensic capacity, organisational cybersecurity controls, international cooperation, platform responsibility and rights-sensitive AI regulation. The study contributes to legal research by showing that AI-driven cybercrime is not only a technical problem but also a legal, institutional and governance challenge.*

Keywords: Artificial intelligence, cybercrime, phishing, deepfakes, automated attacks, digital security governance, cybersecurity law.

INTRODUCTION

Artificial intelligence has changed the nature, scale and legal complexity of cybercrime. Traditionally, cybercrime was understood mainly as the unlawful use of computers, networks or digital systems to commit offences such as unauthorised access, data interference, computer-related fraud, identity theft and online extortion. However, the rapid diffusion of generative AI, machine-learning systems, synthetic media tools and automated attack capabilities has created a new category of technologically

intensified cybercrime: artificial intelligence–driven cybercrime. This refers to cybercriminal conduct in which AI is used either as the main instrument of offending or as a force multiplier that improves the speed, realism, targeting, concealment and scalability of criminal activity. Unlike conventional cyber offending, AI-enabled cybercrime does not merely exploit digital systems; it exploits human trust, institutional weaknesses, identity verification systems and the limits of existing legal and regulatory frameworks.

The importance of this subject is reinforced by recent cyber threat assessments. Europol has warned that criminal groups are using AI to automate attacks, improve social engineering and bypass security measures, thereby making cyberattacks more scalable and efficient.[1] Similarly, ENISA's 2024 threat landscape identifies the growing sophistication of AI-driven phishing and deepfake campaigns, while also noting the persistence of business email compromise and related social-engineering threats.[2] These developments show that AI is not simply a neutral technological innovation. It is increasingly part of the criminal infrastructure through which offenders deceive victims, impersonate trusted persons, generate malicious content and conduct attacks across multiple jurisdictions. In legal terms, this produces a serious governance challenge because cybercrime law, evidence law, data protection law, platform regulation and AI governance must now respond to threats that are fast-moving, transnational and technically complex.

One of the most visible forms of AI-driven cybercrime is AI-enabled phishing. Phishing has long involved deceptive emails, messages or websites designed to trick users into disclosing passwords, financial information or other sensitive data. What AI changes is the quality and scale of deception. Generative AI tools can produce grammatically accurate, context-specific and highly personalised messages in multiple languages. Criminals can use scraped personal data, workplace information or social media activity to craft messages that appear to come from banks, employers, government agencies, delivery companies or professional contacts. The result is a shift from crude mass phishing to automated spear-phishing, where deception is tailored to the victim's role, behaviour and relationships. The FBI's 2024 Internet Crime Report indicated that phishing/spoofing was among the top three cybercrimes by number of complaints reported by victims in 2024, demonstrating the continuing centrality of social engineering in the cybercrime ecosystem.[3] AI therefore worsens an already significant threat by reducing the skill and time required to produce convincing fraudulent communications.

A second major threat is the rise of deepfakes and synthetic identity fraud. Deepfakes involve AI-generated or AI-manipulated audio, images or video that imitate real persons or create false but realistic identities. In cybercrime, deepfakes may be used to impersonate company executives, public officials, family members, lawyers, bank officers or romantic partners. They may also be used to defeat remote identity verification processes, manipulate biometric authentication, blackmail victims or spread false information for financial or political purposes. Europol's 2025 Internet Organised Crime Threat Assessment notes that stolen data can be weaponised in AI-enabled attacks, including the generation of deepfakes, synthetic media and false identities.[4] This creates a serious legal problem because many

[1] Europol, *The Changing DNA of Serious and Organised Crime: EU-SOCTA 2025* (Europol 2025), noting the use of AI for attack automation, social engineering and bypassing security measures.
[2] European Union Agency for Cybersecurity, *ENISA Threat Landscape 2024* (ENISA 2024), identifying AI-driven phishing and deepfake campaigns as growing in sophistication.
[3] Federal Bureau of Investigation, 'FBI Releases Annual Internet Crime Report' (23 April 2025), reporting that phishing/spoofing, extortion and personal data breaches were the top three cybercrimes by number of complaints in 2024.
[4] Europol, *Steal, Deal and Repeat: Internet Organised Crime Threat Assessment 2025* (Europol 2025), discussing the weaponisation of stolen data in AI-enabled attacks, including deepfakes, synthetic media and false identities.

legal systems still depend on documentary identity, voice confirmation, facial recognition, digital signatures and institutional trust. When AI can manufacture believable identity signals, the evidential value of appearance, voice and online presence becomes weaker. This has implications for fraud prevention, banking compliance, criminal evidence, electoral integrity, national security and the protection of vulnerable persons.

A third dimension is the growth of automated cyberattacks. AI can assist attackers in scanning for vulnerabilities, generating malicious code, adapting malware behaviour, evading detection and selecting high-value targets. Although AI is also used defensively for intrusion detection, anomaly monitoring and threat intelligence, the same technological capabilities can be misused by offenders. The concern is not only that AI can create new offences, but that it can industrialise existing ones. Automated attack systems may allow criminals to test thousands of targets, refine attack methods and deploy malware or credential-harvesting tools with limited human supervision. This weakens the traditional assumption that serious cybercrime requires a high level of technical expertise. It also supports the growth of cybercrime-as-a-service markets, where criminal tools, stolen credentials, phishing kits, malware services and synthetic identities can be bought or rented by less technically skilled offenders.

These developments have significant implications for digital security governance. Digital security governance refers to the legal, institutional, technical and policy mechanisms used to protect digital systems, regulate online conduct, investigate cyber offences, allocate responsibility and manage cyber risk. It is broader than cybersecurity in the narrow technical sense. It includes criminal law, regulatory enforcement, corporate compliance, platform accountability, AI transparency rules, data protection duties, incident reporting obligations, international cooperation and digital literacy. AI-driven cybercrime challenges this governance architecture in at least four ways. First, it increases the speed and volume of attacks, making purely reactive enforcement inadequate. Second, it blurs the distinction between real and synthetic evidence, complicating attribution and proof. Third, it creates cross-border harms, as offenders, victims, servers, platforms and payment systems may be located in different jurisdictions. Fourth, it places new responsibilities on private actors, including AI developers, social media platforms, banks, telecom providers, cloud companies and cybersecurity firms.

Existing legal frameworks provide an important but incomplete foundation. The Council of Europe's Budapest Convention on Cybercrime remains a key international instrument for harmonising cybercrime offences and strengthening cooperation among states.[5] The United Nations Convention against Cybercrime, adopted by the UN General Assembly on 24 December 2024, further reflects the global recognition that cybercrime requires international legal coordination and mechanisms for sharing electronic evidence.[6] However, AI-driven cybercrime exposes gaps in frameworks that were largely designed around unauthorised access, data interference, fraud and computer misuse. These laws may criminalise many AI-enabled acts, but they do not always address the specific risks of algorithmic automation, synthetic media, AI-generated impersonation, malicious model use, content provenance or the accountability of AI service providers whose tools may be abused by criminals.

The European Union's Artificial Intelligence Act offers one example of a more risk-based regulatory response. Its transparency obligations require certain disclosures where users interact with AI systems

[5] Council of Europe, *Convention on Cybercrime* ETS No 185 (Budapest, 23 November 2001); Council of Europe, 'The Budapest Convention' describing the Convention as a framework for cooperation among parties in cybercrime cases.

[6] United Nations Office on Drugs and Crime, 'United Nations Convention against Cybercrime', noting that the Convention was adopted by the UN General Assembly on 24 December 2024 by Resolution 79/243.

or where AI systems generate synthetic content, including deepfakes.[7] This is significant because deepfake-related cybercrime thrives where victims cannot distinguish authentic communications from manipulated content. Nevertheless, transparency rules alone cannot solve the problem. A malicious actor who uses AI to commit fraud is unlikely to comply voluntarily with labelling or disclosure duties. For that reason, AI governance must be linked to criminal enforcement, platform moderation, identity verification standards, watermarking technologies, corporate cybersecurity duties and public awareness. The legal challenge is to design rules that deter misuse without suppressing legitimate innovation, research, satire, journalism, cybersecurity testing or lawful AI development.

The governance problem is also complicated by questions of attribution and liability. In conventional cybercrime, investigators already struggle to identify offenders due to anonymisation tools, proxy infrastructure, botnets, cryptocurrency payments and cross-border hosting. AI adds further complexity by enabling synthetic personas, automated decision-making and realistic impersonation. Where an AI system generates a phishing email, clones a voice or assists in vulnerability discovery, legal responsibility may involve several actors: the direct offender, the person who procured the tool, the platform that hosted the content, the developer of the model, the organisation that failed to implement adequate controls, or the intermediary that enabled payment or distribution. A central issue for legal research is therefore how to allocate responsibility fairly and effectively without imposing unrealistic duties on every participant in the digital ecosystem.

This article proceeds from the argument that AI-driven cybercrime requires a shift from a purely offence-based model of cybercrime control to a governance-based model. Criminalisation remains essential, but it is not sufficient. Effective response requires prevention, resilience, detection, cooperation, accountability and rights protection. Frameworks such as the NIST AI Risk Management Framework emphasise the need to manage AI risks to individuals, organisations and society through structured governance and risk management.[8] CISA's AI Roadmap similarly recognises the need to promote beneficial uses of AI for cybersecurity, protect AI systems from cyber threats and deter malicious use of AI capabilities against critical infrastructure.[9] These approaches show that digital security governance must combine legal rules with operational risk management. The article therefore examines AI-enabled phishing, deepfakes and automated attacks as emerging threats to digital security governance. It argues that these threats are not merely technical problems for cybersecurity professionals; they are legal and institutional problems that affect criminal justice, corporate regulation, human rights, privacy, evidence, consumer protection and national security.

Conceptualising Artificial Intelligence–Driven Cybercrime

The literature on cybercrime has traditionally treated digital offending as the misuse of computers, networks, data systems or online platforms for criminal purposes. Early cybercrime scholarship focused on offences such as unauthorised access, malware distribution, identity theft, computer-related fraud, data interference, system interference and cyber-enabled financial crime. These offences remain central to contemporary cybercrime law, especially under instruments such as the Budapest Convention on Cybercrime, which identifies offences including illegal access, illegal interception, data interference,

[7] European Union, 'AI Act: Regulatory Framework for Artificial Intelligence'; see also Article 50 on transparency obligations for AI systems and AI-generated synthetic content.

[8] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1, 2023), describing AI risk management as a means of addressing risks to individuals, organisations and society.

[9] Cybersecurity and Infrastructure Security Agency, 'CISA Releases Roadmap for Artificial Intelligence Adoption' (14 November 2023), outlining efforts to use AI for cybersecurity, protect AI systems and deter malicious use of AI capabilities.

system interference, misuse of devices, computer-related forgery and computer-related fraud.[10] However, the rise of artificial intelligence has expanded the character of cybercrime from computer misuse into automated and intelligent digital manipulation.

Artificial intelligence–driven cybercrime may be understood as criminal conduct in which AI systems are used to plan, facilitate, automate, conceal or scale unlawful activity. Unlike ordinary cybercrime, AI-driven cybercrime does not only depend on technical intrusion into digital systems. It also exploits language, identity, perception, trust, social behaviour and institutional verification processes. Brundage and others argue that AI changes the threat landscape because it can expand existing threats, introduce new threats and alter the typical relationship between attacker capability and defensive capacity.[11] This is important because AI reduces the level of human skill required to carry out sophisticated attacks. A person who lacks advanced coding or linguistic ability may use generative AI to produce convincing phishing emails, fake websites, malicious scripts or deepfake content.

The concept is also closely connected to the dual-use nature of AI. AI tools can support legitimate cybersecurity functions such as threat detection, anomaly analysis, malware classification, vulnerability scanning and automated response. At the same time, the same capabilities can be misused by criminals to automate reconnaissance, produce deception, impersonate individuals and evade detection. The literature therefore presents AI as both a defensive asset and an offensive risk. NIST's AI Risk Management Framework recognises AI as a socio-technical technology whose risks arise not only from technical design but also from the context in which systems are developed, deployed and used.[12] This view is useful for legal research because it shows that AI-driven cybercrime cannot be addressed only by criminal law. It requires governance across technology design, platform responsibility, data protection, digital identity, cybersecurity compliance and law enforcement cooperation.

Cybercrime agencies have increasingly recognised AI as a force multiplier for organised crime. Europol's 2025 Internet Organised Crime Threat Assessment notes that AI-driven techniques may facilitate data acquisition and that stolen data can be weaponised to generate deepfakes, synthetic media and false identities.[13] This means that stolen data is no longer only a commodity for resale; it is also raw material for AI-enabled impersonation and targeted manipulation. Similarly, the broader organised crime literature now links AI to the professionalisation of cybercrime-as-a-service, where criminal services such as phishing kits, malware, credential theft, ransomware tools and identity documents are offered in specialised online markets. In this context, AI does not replace cybercriminal networks; rather, it strengthens them by increasing their operational efficiency and lowering entry barriers for less skilled offenders.

From a legal perspective, AI-driven cybercrime raises difficult questions about attribution, liability, evidence and prevention. Criminal law is usually designed to punish human conduct based on actus reus and mens rea. However, AI-enabled offending often involves layered conduct: one person may create the AI tool, another may deploy it, another may provide stolen data, another may host the infrastructure, and another may receive criminal proceeds. The use of automation also complicates proof of intention, especially where tools are designed for lawful purposes but later repurposed for criminal activity. Therefore, literature on AI-driven cybercrime increasingly calls for a governance

[10] Council of Europe, Convention on Cybercrime ETS No 185 (Budapest, 23 November 2001), arts 2–8; see also Council of Europe, Explanatory Report to the Convention on Cybercrime ETS No 185, para 18.
[11] Miles Brundage and others, The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation (Future of Humanity Institute, University of Oxford and others, 2018).
[12] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1, 2023).
[13] Europol, Steal, Deal and Repeat: Internet Organised Crime Threat Assessment 2025 (Europol 2025).

model that combines criminalisation with risk management, due diligence, technical safeguards, cross-border cooperation and corporate accountability.

AI-Enabled Phishing and Social Engineering

Phishing remains one of the most persistent forms of cybercrime, and the literature shows that AI has made it more dangerous. Traditional phishing often relied on poorly written emails, generic messages and obvious spelling or formatting errors. Many users were trained to identify phishing by looking for suspicious grammar, unusual sender addresses and unrealistic claims. However, generative AI weakens this form of user awareness because it can produce fluent, grammatically accurate and contextually believable messages. This changes the nature of phishing from crude mass deception to personalised social engineering.

The FBI reported that phishing/spoofing was among the top three cybercrimes by number of victim complaints in 2024.[14] This confirms that phishing remains a major threat even before considering the accelerating effects of AI. AI-enabled phishing improves traditional phishing in at least four ways. First, it improves language quality by producing professional and natural messages. Second, it allows multilingual phishing, enabling offenders to target victims across jurisdictions. Third, it allows personalisation by using data from social media, breached databases or corporate websites. Fourth, it supports automation, allowing criminals to generate large numbers of unique messages while avoiding repeated templates that security systems can easily detect.

Hazell's study on spear phishing with Large Language Models (LLMs) shows that LLMs can assist in the reconnaissance and message-generation stages of spear-phishing attacks.[15] This is significant because spear phishing is more dangerous than general phishing. It does not target random users with generic messages; it targets specific individuals or organisations based on their identity, role, workplace, relationships or online behaviour. AI can search publicly available information, infer likely communication styles and produce messages that look like internal workplace communication. For example, a finance officer may receive an AI-generated email that appears to come from a managing director requesting urgent payment approval. A student may receive a fake university portal message. A bank customer may receive a convincing fraud alert that leads to a fake login page.

Schmitt's work on generative AI and social engineering identifies three main areas where generative AI amplifies social engineering: realistic content creation, advanced targeting and personalisation, and automated attack infrastructure.[16] These three pillars are useful for understanding why AI-enabled phishing is not merely a technical problem. It is a behavioural and institutional problem. The attack succeeds because it imitates normal communication patterns and manipulates trust. In legal terms, this creates difficulties for victim protection and corporate liability. If an employee authorises payment after receiving a realistic AI-generated instruction, questions arise over whether the loss should be treated as individual negligence, corporate control failure, bank responsibility or criminal fraud.

The literature also shows that AI-enabled phishing interacts with business email compromise. ENISA's 2024 threat landscape identifies a sharp increase in business email compromise campaigns, AI-powered phishing campaigns and deepfake-based scams targeting executives and high-value individuals.[17] This is important for digital security governance because BEC attacks often bypass purely technical

[14] Federal Bureau of Investigation, 'FBI Releases Annual Internet Crime Report' (23 April 2025).
[15] Julian Hazell, *Spear Phishing with Large Language Models* (2023) arXiv preprint arXiv: 2305.06972
[16] Michael Schmitt, 'Digital Deception: Generative Artificial Intelligence in Social Engineering Attacks' (2024) *Artificial Intelligence Review*.
[17] European Union Agency for Cybersecurity, *ENISA Threat Landscape 2024* (ENISA 2024).

defences. They do not always require malware. Instead, they exploit authority, urgency and routine administrative procedures. AI strengthens these attacks by creating emails that match professional tone and organisational context.

Another issue in the literature is the use of adversary-in-the-middle phishing tools, which can bypass multi-factor authentication by intercepting credentials and session tokens.[18] This shows that AI-enabled phishing should not be treated simply as "fake email." It increasingly forms part of a broader attack chain involving fake websites, credential harvesting, session hijacking, payment diversion and identity theft. Therefore, effective governance requires stronger authentication, phishing-resistant MFA, staff training, incident reporting, domain monitoring and liability rules that encourage institutions to prevent foreseeable deception.

Deepfakes, Synthetic Identity and Digital Impersonation

Deepfakes represent one of the most legally disruptive forms of AI-driven cybercrime. Deepfakes are synthetic or manipulated audio, video or image content generated by AI to make a person appear to say or do something they did not say or do. Mirsky and Lee explain that deepfake technology has advanced to the point where it is increasingly difficult to distinguish real content from fake content, especially where the audience lacks technical detection tools.[19] While deepfakes may be used for entertainment, satire or creative expression, their criminal misuse includes fraud, extortion, identity theft, reputational harm, election manipulation, harassment and blackmail.

Chesney and Citron argue that deepfakes create serious risks for privacy, democracy and national security because they make false audio-visual evidence more believable and can be used for exploitation, intimidation and sabotage.[20] Their argument remains highly relevant to AI-driven cybercrime because deepfakes attack the evidential foundation of trust. In ordinary communication, people often rely on voice, face and video as proof of identity. Where these identity signals can be artificially manufactured, individuals and organisations become vulnerable to impersonation.

In financial crime, deepfakes may be used to impersonate executives, clients, bank officers or family members. Voice cloning may be used to instruct employees to make transfers, approve invoices or disclose confidential information. Video deepfakes may be used to pass remote identity checks, deceive business partners or manipulate online meetings. Europol's 2025 serious and organised crime assessment warns that AI-powered voice cloning and live video deepfakes enable new forms of fraud, extortion and identity theft.[21] This is especially concerning because many organisations now use remote onboarding, virtual meetings and online verification procedures. The more institutions depend on digital identity, the more valuable synthetic identity becomes to criminals.

Deepfake cybercrime also creates evidential challenges for courts and investigators. Traditionally, video or audio evidence has been treated as powerful evidence because it appears to show events directly. Deepfakes weaken that assumption. Courts may increasingly need expert evidence on authenticity, metadata, chain of custody and forensic analysis. At the same time, there is a risk of the "liar's dividend," where genuine evidence is dismissed as fake simply because deepfakes exist. This has serious implications for criminal justice because both false incrimination and false denial may become easier.

[18] ibid.
[19] Yisroel Mirsky and Wenke Lee, 'The Creation and Detection of Deepfakes: A Survey' (2021) 54 *ACM Computing Surveys* 1.
[20] Robert Chesney and Danielle Keats Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 California Law Review 1753.
[21] Europol, *The Changing DNA of Serious and Organised Crime: EU-SOCTA 2025* (Europol 2025).

The legal literature also connects deepfakes to privacy and dignity. A person whose face or voice is used without consent suffers more than economic loss. They may suffer reputational damage, emotional harm, loss of autonomy and violation of personal identity. Where deepfakes are used in sexual exploitation or political manipulation, the harm may extend to democratic participation, public trust and gender-based abuse. Therefore, deepfake governance cannot be limited to cybercrime statutes. It must also involve privacy law, data protection law, harassment law, platform regulation, evidence law and human rights protections.

The EU AI Act attempts to respond partly through transparency obligations. Article 50 creates obligations relating to AI systems that interact with users and systems that generate synthetic content, including deepfakes.[22] These transparency rules are useful because they recognise deception as a regulatory harm. However, the literature also suggests that transparency is insufficient where actors are deliberately criminal. A fraudster who creates a deepfake to steal money will not voluntarily label it as artificial. Therefore, transparency obligations must be supported by technical watermarking, platform detection, criminal penalties, victim remedies, authentication protocols and public education.

Automated Attacks, Malware and Cybercrime-as-a-Service

Another major theme in the literature is the use of AI to automate cyberattacks. Automation is not new in cybercrime. Botnets, worms, credential-stuffing tools and automated scanners existed before modern generative AI. However, AI changes automation by making it more adaptive, intelligent and accessible. AI systems may assist attackers in identifying vulnerabilities, writing malicious code, modifying malware, generating fake content, testing defences and selecting targets. This does not mean that AI automatically creates highly sophisticated malware without human direction. Rather, it means that AI can support different stages of the attack lifecycle.

Brundage and others note that AI may increase the scale and efficiency of cyberattacks by automating tasks that previously required human labour or specialist expertise.[23] This is central to the idea of AI as a force multiplier. A criminal group can use AI to reduce the time needed for reconnaissance, generate attack variations, create convincing lures and adapt communication to victim responses. In ransomware operations, AI may help identify high-value targets, analyse stolen documents and craft pressure messages. In fraud schemes, AI may help maintain long conversations with victims through chatbots, fake customer service channels or romance scams.

The literature on cybercrime-as-a-service is also relevant. Cybercrime has become increasingly commercialised. Criminal actors may buy malware, rent botnets, purchase stolen credentials, subscribe to phishing kits or pay for access to compromised systems. AI can strengthen this ecosystem by enabling service providers to offer more advanced products to a wider market. For example, an offender may not know how to write a phishing email, create a fake website or generate malware code, but may access tools that assist with those tasks. Europol's IOCTA links stolen data, AI-enabled attacks and crime-as-a-service markets, showing that data-driven criminal economies are becoming more integrated.[24]

Automated attacks also create challenges for cybersecurity governance because speed matters. A human-led response may be too slow where attacks are generated, tested and modified automatically. Organisations therefore need automated defensive tools, threat intelligence sharing and continuous monitoring. However, this raises further governance questions. If organisations use AI defensively, they

[22] Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, art 50; European Commission, 'Code of Practice on Marking and Labelling of AI-Generated Content'.
[23] Brundage and others (n 11).
[24] Europol, *Steal, Deal and Repeat* (n 13).

must ensure that these systems are reliable, lawful, proportionate and not discriminatory. A cybersecurity tool that wrongly flags legitimate activity or processes excessive personal data may create privacy and due process concerns. Therefore, AI-driven cybercrime leads to an "AI versus AI" security environment, where both attackers and defenders use automated systems.

The CISA AI Roadmap recognises this dual character of AI by focusing on the beneficial use of AI for cybersecurity, the protection of AI systems and the need to deter malicious use of AI against critical infrastructure.[25] This indicates that digital security governance must treat AI systems themselves as critical assets. If AI models are manipulated, poisoned, stolen or misused, they can become attack surfaces. For example, attackers may attempt prompt injection, data poisoning, model inversion or adversarial attacks against AI-enabled security systems. Thus, the literature increasingly expands from cybercrime committed with AI to cybercrime committed against AI systems.

Digital Security Governance and Regulatory Responses

The literature on digital security governance suggests that AI-driven cybercrime cannot be controlled by criminal law alone. Criminal law is necessary because it defines prohibited conduct and allows investigation, prosecution and punishment. However, cybercrime is often cross-border, fast-moving and difficult to attribute. Many offenders operate outside the victim's jurisdiction. Evidence may be stored in multiple countries. Platforms may be private companies with their own moderation policies. Victims may fail to report attacks due to embarrassment, fear or lack of confidence. These features require governance beyond punishment.

International cooperation is central to cybercrime governance. The Budapest Convention remains a foundational instrument because it harmonises offences and procedural powers and promotes international cooperation.[26] However, the global landscape has expanded with the UN Convention against Cybercrime, adopted by the UN General Assembly on 24 December 2024.[27] The UN Convention is important because it represents a broader global attempt to create a common framework for combating cybercrime and sharing electronic evidence. Nevertheless, international cybercrime instruments face practical challenges. States differ in legal systems, investigative capacity, human rights safeguards, data access rules and political willingness to cooperate.

Cybersecurity regulation is another important governance tool. The EU NIS2 Directive establishes a framework for cybersecurity across critical sectors and requires risk management and incident reporting by essential and important entities.[28] Such obligations are relevant because AI-driven cybercrime often targets organisations rather than only individuals. Banks, healthcare providers, universities, public agencies, energy companies and telecom providers may all face AI-enabled phishing, ransomware, impersonation and automated attacks. Governance therefore requires institutions to adopt preventive security measures, not merely report crimes after loss has occurred.

Platform governance is also significant. AI-driven deception often spreads through social media, messaging platforms, online marketplaces, video platforms and email services. The EU Digital Services Act introduces due diligence obligations for online services, including mechanisms for addressing illegal content and systemic risks.[29] While the DSA is not specifically a cybercrime statute, it is relevant

[25] Cybersecurity and Infrastructure Security Agency, 'CISA Releases Roadmap for Artificial Intelligence Adoption' (14 November 2023).
[26] Council of Europe, Convention on Cybercrime (n 10).
[27] United Nations Office on Drugs and Crime, 'United Nations Convention against Cybercrime', noting adoption by UN General Assembly Resolution 79/243 on 24 December 2024.
[28] European Commission, 'NIS2 Directive: Securing Network and Information Systems' (20 January 2026).
[29] European Commission, 'The Digital Services Act'.

because deepfake scams, impersonation networks, fake accounts and fraudulent advertisements often depend on digital platforms for distribution. A governance approach must therefore ask what duties platforms should have to detect, remove, label or reduce harmful synthetic content.

AI-specific governance is developing alongside cybercrime and cybersecurity law. The EU AI Act uses a risk-based model and includes transparency obligations for certain AI outputs.[30] This approach recognises that AI risks cannot be managed only after harm occurs. Developers and deployers may need to build safeguards into systems before deployment. However, AI regulation faces a major enforcement challenge in cybercrime contexts. Criminals may use open-source models, stolen models, jailbroken systems or foreign-hosted services. Therefore, AI governance must combine transparency, safety-by-design, access control, content provenance, model monitoring and cooperation with law enforcement.
The literature also emphasises the importance of private-sector responsibility. Much digital security infrastructure is controlled by private actors: cloud providers, banks, telecom companies, social media platforms, AI developers, cybersecurity vendors and payment processors. These actors often detect cybercrime before the state does. They also control the systems through which attacks are launched or prevented. As a result, governance increasingly depends on public-private cooperation. However, this cooperation must be balanced with rights protections. Overbroad monitoring may threaten privacy. Excessive content removal may affect freedom of expression. Automated risk scoring may create discrimination. Therefore, digital security governance must be both security-focused and rights-sensitive.

Gap in Knowledge

The existing literature provides strong evidence that AI intensifies phishing, deepfakes and automated cyberattacks. However, several gaps remain. First, much of the literature focuses on technical risk, while fewer studies examine how existing criminal law doctrines can be adapted to AI-enabled offending. More legal analysis is needed on intention, complicity, negligence, corporate liability and platform responsibility in cases where AI tools are used indirectly.

Second, there is still limited empirical evidence on the actual scale of AI-driven cybercrime. Many reports identify growing risks, but it is often difficult to separate confirmed AI-enabled attacks from ordinary cybercrime that merely appears sophisticated. This creates a measurement problem. Policymakers may overstate or understate the AI threat depending on limited data. Better reporting frameworks are needed to identify whether AI was used in phishing, fraud, malware development or synthetic identity crime.

Third, the literature has not fully resolved the tension between AI innovation and AI restriction. Strong regulation may reduce misuse but may also limit legitimate cybersecurity research, open-source development and beneficial AI applications. Weak regulation may encourage innovation but leave society vulnerable to large-scale deception and automated crime. This tension is especially difficult because AI tools are general-purpose technologies. The same model can draft a lawful business email or a phishing email; it can generate educational content or fraudulent content.

Fourth, further research is needed on the evidential implications of deepfakes. Courts, investigators and lawyers need clearer standards for authenticating digital evidence in an environment where audio and video can be manipulated. The literature identifies the problem, but legal systems still need practical rules on admissibility, expert evidence, burden of proof and digital forensic standards.

[30] Regulation (EU) 2024/1689, art 50; European Commission, 'Code of Practice on Marking and Labelling of AI-Generated Content' (n 22).

Finally, there is a governance gap between international cybercrime law and AI regulation. Cybercrime instruments focus on offences, procedural powers and cooperation. AI regulation focuses on risk, transparency and system design. Yet AI-driven cybercrime sits between both fields. A stronger literature should therefore integrate cybercrime law, AI governance, cybersecurity regulation, digital evidence and human rights. This article contributes to that gap by treating AI-driven cybercrime not only as a technological threat but as a challenge to digital security governance.

Implications of the Study

This study shows that artificial intelligence–driven cybercrime has important implications for law, digital security governance and institutional protection. AI has made cybercrime more sophisticated by improving phishing messages, enabling deepfake impersonation and supporting automated attacks. Therefore, cybercrime should no longer be treated only as a technical problem, but also as a legal and governance issue.

First, the study implies that existing cybercrime laws need to be updated or interpreted more broadly to address AI-enabled offences. Traditional laws on fraud, identity theft, unauthorised access and data misuse may apply to many AI-related crimes, but they may not fully cover issues such as deepfake impersonation, synthetic identities and AI-generated deception. This means lawmakers must consider clearer rules on liability, especially where AI tools are used to assist or automate criminal conduct.

Second, the study highlights the need for stronger digital evidence procedures. Deepfakes make it harder to trust audio, video and images as proof. Courts, investigators and law enforcement agencies may therefore need better forensic tools, authentication standards and digital evidence rules to confirm whether media content is real or AI-generated.

Third, the study implies that organisations must improve their cybersecurity governance. Businesses, banks, schools, hospitals and government agencies should not rely only on passwords, emails or voice confirmation for sensitive actions. They need stronger verification systems, staff training, phishing-resistant authentication and clear procedures for reporting suspicious activities.

Finally, the study shows that AI developers, platforms and regulators must share responsibility for reducing harmful AI misuse. Transparency rules, content detection, watermarking and platform monitoring can help reduce risks, although they cannot completely stop criminal misuse. Overall, the implication is that digital security governance must become more preventive, coordinated and rights-sensitive in response to AI-driven cybercrime.

CONCLUSION

This article has examined artificial intelligence–driven cybercrime as an emerging threat to digital security governance. It focused on three major forms of AI-enabled cybercrime: phishing, deepfakes and automated attacks. The discussion showed that AI does not merely create new cyber threats; it strengthens existing forms of cybercrime by making them faster, more realistic, more scalable and more difficult to detect. AI-enabled phishing allows criminals to produce personalised and convincing messages. Deepfakes weaken trust in voice, image and video evidence. Automated attacks increase the speed and efficiency of cybercriminal operations.

The main argument of the article is that AI-driven cybercrime requires a governance response that goes beyond traditional criminal law. Criminalisation remains necessary, but it is not sufficient. Effective response requires a combination of legal reform, cybersecurity regulation, institutional risk management, international cooperation, platform responsibility, AI transparency and public awareness.

This is because AI-driven cybercrime operates across legal, technical and social boundaries. It affects criminal justice, corporate governance, data protection, digital identity, evidence law and human rights. The study also shows that digital security governance must become more adaptive. Legal systems designed for earlier forms of computer misuse may struggle to address AI-generated impersonation, synthetic identities and automated deception. Therefore, governments and regulators must strengthen cybercrime laws while also ensuring that such laws remain compatible with due process, privacy and freedom of expression. Organisations must also improve their internal controls, particularly in relation to identity verification, financial authorisation, employee training and incident response.

In conclusion, AI-driven cybercrime is one of the most serious emerging challenges in the digital environment. Its danger lies not only in the technology itself, but in its ability to manipulate trust, automate deception and weaken confidence in digital evidence. A strong governance response must therefore be preventive, cooperative, rights-sensitive and technologically informed. The future of digital security will depend on how effectively law, policy and institutions respond to the criminal misuse of artificial intelligence.