Indonesian Journal *of* Digital Crime *and* Criminal Justice
Vol. 1 No. 1. January 2025. pp. 125-162
E-ISSN: XXXX-XXXX| P-ISSN: XXXX-XXX
Original Research Article

Generative Artificial Intelligence and Cybercrime: Emerging Threats and Regulatory Responses

Zainuddin Ahmad Surya [1]*
[1] Universitas Islam Negeri Syarif Hidayatullah Jakarta, Tangerang Selatan, Indonesia
* Corresponding author: zainuddin@gmail.com

ABSTRACT

Generative artificial intelligence (GenAI) has introduced transformative capabilities in content creation, but it also presents new opportunities for cybercriminal activities. The ability of GenAI systems to produce realistic text, images, audio, and code has facilitated the emergence of sophisticated cyber fraud, phishing schemes, identity theft, and automated hacking tools. This article investigates the evolving relationship between generative AI and cybercrime, focusing on emerging threats and regulatory responses at both national and international levels. Using doctrinal legal analysis and comparative evaluation, the study examines existing cybercrime regulations and their adequacy in addressing AI-enabled offences. The findings reveal that current legal frameworks are largely reactive and insufficient to address the speed and scale of GenAI-driven criminal innovation. Key challenges include attribution of responsibility, detection of AI-generated content, and jurisdictional complexities in cross-border digital crimes. The article proposes a multi-layered regulatory approach combining criminal law reform, platform accountability, and AI governance frameworks. It concludes that effective regulation must anticipate technological evolution rather than respond retrospectively, ensuring that criminal justice systems remain resilient in the face of rapidly advancing generative technologies.

KEYWORDS: Generative AI, Cybercrime, Digital Fraud, Artificial Intelligence Regulation, Criminal Law, Online Security

INTRODUCTION

The rapid development of Generative Artificial Intelligence (GenAI) represents a paradigm shift in the architecture of contemporary digital

Available online at https://journal.criminallawinstitute.org/JDCCJ/article/view/19, since January 15, 2025
Submitted: November 11, 2024 ■ Reviewed: November 27, 2024 ■ Revised: January 5, 2026 ■ Accepted: January 15, 2025

ecosystems. Unlike traditional analytical AI systems that are designed primarily to classify, predict, or optimize existing datasets, GenAI systems are capable of producing novel outputs that closely resemble human-generated content. These systems include large language models (LLMs), diffusion models for image generation, and advanced voice and video synthesis technologies, all of which have significantly expanded the boundaries of machine capability (Floridi and Chiriatti 2020).

Large Language Models such as GPT-based architectures operate through deep neural networks trained on vast corpora of textual data, enabling them to generate coherent and contextually relevant text. Similarly, diffusion models have revolutionized image generation by iteratively refining random noise into highly realistic visual outputs. In parallel, voice and video synthesis technologies enable the replication of human speech patterns, facial expressions, and behavioral characteristics with increasing accuracy. Collectively, these innovations represent a shift from AI as a purely analytical tool toward AI as a generative and creative system.

This transformation reflects a broader evolution from analytical AI to generative AI. While earlier AI systems were primarily designed to assist human decision-making through classification and prediction, GenAI systems are capable of autonomous content creation. This includes text generation, code synthesis, image and video production, and even simulation of human interaction. Such capabilities significantly expand the scope of AI applications across industries, including education, healthcare, entertainment, and law enforcement.

GenAI systems also introduce four core functional capabilities relevant to cybercrime analysis. First, content generation enables automated production of text, images, audio, and video at scale. Second, automation allows repetitive tasks to be executed without human intervention, significantly increasing efficiency and scalability. Third, personalization enables systems to tailor outputs based on user behavior or preferences, enhancing persuasive capabilities. Fourth, imitation—including deepfake generation—allows the replication of identities, voices, and visual appearances with high fidelity. Among these, imitation capabilities are

particularly significant in the context of cybercrime due to their potential for deception, fraud, and identity manipulation (Chesney and Citron 2019).

Cybercrime has undergone a significant evolutionary trajectory alongside technological development. Early forms of cybercrime were primarily characterized by manual hacking activities, basic malware deployment, and unauthorized access to computer systems. These offenses typically required specialized technical skills and were limited in scale due to the manual nature of execution. The second generation of cybercrime emerged with the development of scripting tools and automation techniques. During this phase, attackers increasingly relied on pre-written scripts, exploit kits, and automated scanning tools to identify system vulnerabilities. This reduced the technical barrier to entry and enabled a broader range of actors to engage in cybercriminal activities.

A third phase of cybercrime evolution is characterized by platform-based criminal economies. The emergence of ransomware-as-a-service (RaaS), malware-as-a-service, and dark web marketplaces has transformed cybercrime into a highly organized and commodified ecosystem. In this model, technical specialists develop tools and infrastructure, while less skilled actors purchase access to deploy attacks (Europol 2022).

The most recent phase involves the integration of artificial intelligence into cybercrime operations. AI-enabled cybercrime represents a qualitative transformation rather than a mere incremental development. Generative AI systems enable attackers to automate phishing campaigns, generate highly convincing social engineering content, create deepfake identities, and develop adaptive malware. This shift fundamentally alters the structure, scale, and sophistication of cybercrime activities, making them more accessible, scalable, and difficult to detect.

The emergence of GenAI introduces several critical challenges for contemporary cybersecurity governance. First, it significantly lowers the skill barrier required to commit cybercrime. Tasks that previously required advanced technical expertise, such as crafting phishing emails or producing fraudulent digital identities, can now be executed using simple natural language prompts. This democratization of offensive cyber capabilities

increases the number of potential threat actors and expands the overall risk landscape.

Second, GenAI dramatically increases the scale and speed of cybercriminal activities. Automated content generation allows attackers to produce thousands of phishing messages, fake profiles, or malicious scripts within seconds. This scalability fundamentally alters the economics of cybercrime by reducing costs and increasing potential returns. Third, existing regulatory frameworks remain largely based on a human-centered understanding of criminal conduct. Traditional cybercrime laws assume identifiable human perpetrators engaging in discrete acts of wrongdoing. However, GenAI introduces semi-autonomous systems capable of generating harmful content with minimal human intervention, thereby complicating attribution, intent analysis, and legal responsibility.

Existing literature on artificial intelligence and cybersecurity has developed along several distinct but fragmented strands. First, AI ethics scholarship—particularly from institutions such as the OECD and UNESCO—has focused on normative principles such as fairness, transparency, and accountability in AI development (OECD 2019; UNESCO 2021). Second, cybersecurity law literature, including frameworks such as the Budapest Convention on Cybercrime, has addressed cross-border cooperation and procedural standards for cybercrime enforcement. Third, AI governance literature has explored regulatory models for managing algorithmic systems across sectors.

Despite these contributions, significant gaps remain. Existing scholarship has not sufficiently addressed the structural transformation of cybercrime caused by GenAI technologies. Most analyses continue to conceptualize cybercrime as a human-driven activity augmented by technology, rather than as a phenomenon fundamentally reshaped by generative systems. Moreover, the concept of AI as a "crime multiplier"—a force that amplifies the scale, efficiency, and accessibility of criminal conduct—remains under-theorized in legal scholarship. Finally, traditional criminal law doctrines, which rely heavily on human intent, causation, and agency, have not been adequately reassessed in light of GenAI-enabled offenses.

This article is guided by three central research questions:

1. How does generative artificial intelligence transform the structure and characteristics of cybercrime?
2. Why are existing cybercrime regulatory frameworks increasingly inadequate in addressing GenAI-enabled threats?
3. What regulatory models are required to effectively govern and mitigate risks associated with GenAI-driven cybercrime?

This article advances the central argument that Generative Artificial Intelligence does not merely introduce new categories of cybercrime but fundamentally transforms the nature of cybercriminal activity itself. GenAI functions as a crime multiplier that enhances automation, scalability, and transnational reach, thereby altering the structural conditions under which cybercrime occurs. As a result, existing regulatory frameworks—designed for human-centered criminal behavior—are increasingly insufficient. Addressing these challenges requires a shift toward risk-based, adaptive, and multilayered governance models that integrate technological oversight, international cooperation, and redefined principles of criminal responsibility.

CONCEPTUAL FOUNDATIONS: GENERATIVE AI AND CYBERCRIME TRANSFORMATION

A. Concept of Generative Artificial Intelligence

Generative Artificial Intelligence (GenAI) refers to a class of machine learning systems capable of producing novel outputs—such as text, images, audio, video, and software code—based on patterns learned from large-scale datasets. Unlike traditional artificial intelligence systems that primarily perform classification, prediction, or optimization tasks, GenAI systems are designed to generate content that mimics human creativity and communication (Russell and Norvig 2021; Goodfellow, Bengio, and Courville 2016).

One of the most prominent forms of GenAI is the Large Language Model (LLM). LLMs, such as transformer-based architectures, are trained on massive textual corpora and are capable of generating coherent, contextually

relevant, and linguistically sophisticated text. These systems operate through probabilistic modeling of language sequences, enabling them to perform tasks such as drafting documents, answering questions, and simulating conversational interactions (Brown et al. 2020). In the context of cybercrime, LLMs can be misused to generate highly convincing phishing emails, social engineering scripts, and fraudulent communications.

Diffusion models represent another major category of generative systems, particularly in the domain of image and video synthesis. These models work by gradually transforming random noise into structured outputs through iterative refinement processes. Applications such as DALL·E and Stable Diffusion demonstrate the capacity of such systems to generate highly realistic visual content, including synthetic faces, environments, and manipulated scenes (Ho et al. 2020). In cybercriminal contexts, diffusion models can facilitate identity fabrication, deepfake imagery, and visual deception campaigns.

Voice cloning systems constitute a third category of GenAI technologies. These systems use deep neural networks to replicate human voice characteristics, including tone, pitch, accent, and speech patterns. By analyzing short audio samples, voice synthesis models can generate highly realistic speech outputs that are often indistinguishable from the original speaker (Jia et al. 2018). This capability has significant implications for fraud, impersonation, and deception-based cybercrime. Finally, code generation systems represent an emerging dimension of GenAI applications. These systems can produce functional software code based on natural language prompts, enabling users to automate programming tasks. While beneficial for software development, such systems may also be exploited to generate malicious scripts, automate exploit development, or design malware variants (Chen et al. 2021). Collectively, these four categories illustrate the broad technological foundation of GenAI and its dual-use potential.

B. Functional Capabilities Relevant to Cybercrime

The relevance of Generative AI to cybercrime is best understood through its core functional capabilities, which significantly expand the operational

capacity of malicious actors.

(a) Automation Capability

Automation is one of the most transformative features of GenAI in cybercrime contexts. Traditionally, cyberattacks required substantial human involvement in crafting messages, identifying targets, and executing operations. With GenAI, many of these tasks can now be automated. For instance, AI systems can generate phishing emails at scale, simulate human conversations for social engineering attacks, or autonomously adapt malware scripts based on system responses (Brundage et al. 2018). This automation reduces the time, cost, and expertise required to conduct cyberattacks, thereby increasing their frequency and accessibility.

(b) Scalability

Scalability refers to the ability of GenAI systems to generate vast quantities of content with minimal incremental cost. In cybercrime, this translates into mass-scale attacks such as bulk phishing campaigns, automated account creation, and large-scale misinformation dissemination. Unlike manual cyber operations, which are limited by human capacity, GenAI enables near-infinite replication of malicious content. This scalability fundamentally alters the economics of cybercrime by increasing the potential impact of individual attackers (Europol 2022).

(c) Hyper-personalization

Hyper-personalization is another critical capability of GenAI, enabling attackers to tailor malicious content to specific individuals or groups. By analyzing publicly available data or stolen datasets, GenAI systems can generate highly targeted messages that reflect the language, interests, or social context of victims. This significantly increases the effectiveness of social engineering attacks, as personalized content is more likely to deceive recipients (Floridi 2014). Hyper-personalization thus transforms cybercrime from generic mass attacks into precision-targeted operations.

(d) Realism

The realism of GenAI-generated content represents one of its most concerning features. Advances in deep learning have made it possible to generate text, images, audio, and video that are nearly indistinguishable from authentic human-generated content. Deepfake technologies, in particular, pose significant risks by enabling the creation of synthetic identities, fabricated evidence, and manipulated audiovisual recordings (Chesney and Citron 2019). High levels of realism reduce the ability of individuals and institutions to detect fraud, thereby increasing the success rate of cybercriminal activities.

C. Concept of Cybercrime (Doctrinal Perspective)

From a doctrinal legal perspective, cybercrime refers to unlawful activities conducted through or targeting computer systems, digital networks, or information technologies. Classical definitions of cybercrime emphasize the use of digital tools as either the instrument or object of criminal conduct (Wall 2007; Brenner 2010). In this sense, cybercrime encompasses both traditional offenses facilitated by technology and novel offenses that arise uniquely within digital environments.

Cybercrime is commonly categorized into several core legal elements. The first is unauthorized access, which involves gaining entry into computer systems, networks, or data without permission. This includes hacking, password breaches, and intrusion into protected systems. Unauthorized access undermines the integrity and confidentiality of digital infrastructures.

The second category is fraud, which involves the use of digital systems to deceive individuals or institutions for financial or personal gain. Cyber fraud includes phishing schemes, identity theft, and online financial scams. The rise of GenAI significantly amplifies fraud risks by enabling the creation of highly convincing deceptive content. The third element is data interference, which refers to the unauthorized alteration, deletion, or corruption of digital data. This includes malware attacks, ransomware encryption, and data manipulation. Such conduct compromises the integrity and reliability of information systems.

The fourth element is system interference, which involves disrupting the functioning of computer systems or networks. Distributed denial-of-service (DDoS) attacks and other forms of system overload fall within this category. System interference affects the availability and stability of digital infrastructures. These doctrinal categories, however, were developed in a pre-GenAI era and assume human-centered criminal conduct. As a result, they may not adequately capture the complexities introduced by autonomous or semi-autonomous AI systems that can independently generate harmful outputs.

D. Theoretical Lens: AI-Amplified Crime

To address the limitations of traditional cybercrime frameworks, this article introduces the concept of "AI-amplified cybercrime." This concept refers to criminal activities that are significantly enhanced by artificial intelligence technologies in terms of scale, speed, sophistication, and accessibility. Rather than replacing human actors entirely, GenAI functions as a force multiplier that enhances existing criminal capabilities and reduces operational constraints.

AI-amplified cybercrime differs from conventional cybercrime in several respects. First, it increases the scale of operations by enabling mass production of malicious content. Second, it accelerates the speed of execution by automating key stages of the attack lifecycle. Third, it enhances complexity by enabling adaptive and context-sensitive attacks. Finally, it lowers entry barriers by making advanced cyber capabilities accessible to non-expert users. This conceptualization aligns with broader literature on technological amplification of crime, which suggests that digital tools do not merely facilitate criminal behavior but fundamentally reshape its structure and dynamics (Wall 2007; Brenner 2010). In the context of GenAI, amplification occurs at both the operational and strategic levels, transforming cybercrime into a more scalable and resilient phenomenon.

E. Analytical Framework

This article adopts a three-dimensional analytical framework to examine GenAI-enabled cybercrime. The first dimension is capability expansion, which examines how GenAI enhances the technical and operational capabilities of cybercriminal actors. This includes improvements in content generation, deception techniques, and automated decision-making.

The second dimension is crime automation, which focuses on the extent to which criminal processes can be executed with minimal human intervention. Automation fundamentally alters the role of human agency in cybercrime, shifting from direct execution to supervisory or orchestration roles. The third dimension is regulatory mismatch, which analyzes the gap between existing legal frameworks and the evolving nature of AI-enabled cybercrime. Traditional cybercrime laws are largely designed around identifiable human perpetrators and discrete acts of misconduct. However, GenAI introduces distributed, automated, and scalable forms of criminality that challenge these assumptions. This mismatch creates enforcement difficulties, evidentiary challenges, and jurisdictional complexities. Together, these three dimensions provide a comprehensive framework for understanding the structural transformation of cybercrime in the era of generative artificial intelligence.

EMERGING FORMS OF GENAI-ENABLED CYBERCRIME

The integration of Generative Artificial Intelligence into digital ecosystems has fundamentally reshaped the typology of cybercrime. Unlike traditional cyber threats that relied heavily on human expertise and manual execution, GenAI-enabled cybercrime is characterized by automation, scalability, and adaptive intelligence. This transformation is not merely incremental but structural, as it alters the modus operandi of criminal actors and expands the range of feasible cyber offenses. The following sections examine the most significant emerging forms of GenAI-enabled cybercrime.

A. AI-Powered Phishing and Social Engineering

One of the earliest and most rapidly evolving applications of GenAI in cybercrime is AI-powered phishing and social engineering. Phishing, traditionally defined as the deceptive attempt to obtain sensitive information by impersonating legitimate entities, has become significantly more sophisticated through the use of generative systems.

Spear phishing has been enhanced by AI-driven personalization capabilities. Whereas traditional phishing campaigns relied on generic mass emails, GenAI systems can now analyze publicly available data from social media platforms, professional networks, and leaked datasets to craft highly targeted messages. These messages often mimic the linguistic style, professional context, and behavioral patterns of trusted contacts, thereby increasing their effectiveness (Brundage et al. 2018).

Automated email generation further amplifies this threat. Large Language Models are capable of producing grammatically accurate, contextually relevant, and psychologically persuasive emails at scale. This automation allows attackers to generate thousands of customized phishing messages within seconds, significantly reducing the cost and skill required for cyber deception. The sophistication of these messages makes traditional detection mechanisms, such as keyword filtering and spam detection systems, increasingly ineffective (Chesney and Citron 2019).

Conversational scam bots represent another emerging vector of social engineering. These AI-driven agents can engage victims in real-time conversations, simulate customer support interactions, or impersonate trusted individuals. Unlike static phishing emails, conversational bots adapt dynamically to user responses, increasing their persuasive power and reducing suspicion. This interactive dimension marks a significant evolution in social engineering techniques.

B. Identity Fraud and Deepfake Exploitation

Identity fraud has been significantly transformed by advances in generative AI technologies. Deepfake systems, which utilize neural networks to synthesize realistic audio-visual content, enable the creation of highly

convincing impersonations of individuals.

Voice cloning technology allows attackers to replicate a person's voice using only short audio samples. These synthetic voices can be used to impersonate executives, family members, or public officials in fraudulent communications. The realism of voice synthesis makes it difficult for victims to distinguish between authentic and artificial speech, thereby increasing the success rate of impersonation fraud (Jia et al. 2018).

Synthetic identity creation represents another major development. GenAI systems can generate entirely fictitious identities, including names, photographs, biographies, and digital footprints. These synthetic identities can be used to open financial accounts, bypass verification systems, or conduct fraudulent transactions. Unlike traditional identity theft, which relies on stolen real-world identities, synthetic identity fraud creates entirely new personas that are difficult to trace or verify.

Impersonation fraud is further amplified by multimodal deepfake capabilities. Attackers can combine synthetic voice, video, and text to create highly convincing representations of individuals. Such capabilities are particularly dangerous in contexts involving executive decision-making, legal communications, or financial authorization processes.

C. Financial Cybercrime

GenAI has also significantly expanded the scope of financial cybercrime, particularly through the automation and enhancement of fraud schemes. CEO fraud, also known as business email compromise (BEC), has become increasingly sophisticated through the use of AI-generated communication. Attackers can now replicate the communication style of corporate executives, enabling highly convincing fraudulent instructions to employees or financial departments. The integration of contextual awareness further enhances the credibility of such attacks (Europol 2022).

Investment scams have similarly evolved with the assistance of generative systems. AI-driven platforms can generate persuasive financial narratives, fake testimonials, and simulated investment performance reports. These systems exploit behavioral biases and emotional triggers to manipulate

victims into fraudulent investment schemes. The scalability of GenAI allows such scams to be deployed across multiple jurisdictions simultaneously.

Automated banking fraud represents another emerging threat. AI systems can assist in generating scripts or tools designed to bypass authentication mechanisms, exploit vulnerabilities in banking systems, or automate transaction fraud. While financial institutions employ advanced security systems, the adaptive nature of GenAI-enabled attacks increases the complexity of defense mechanisms.

D. Malware and Exploit Generation

One of the most technically significant developments in cybercrime is the use of GenAI for malware and exploit generation. Traditional malware development required advanced programming skills and deep technical knowledge. However, generative systems have lowered this barrier significantly. AI-generated malware refers to malicious software created or assisted by generative models. These systems can produce code that performs harmful functions such as data exfiltration, system corruption, or unauthorized access. While current models often include safety restrictions, adversarial techniques can circumvent these safeguards, enabling malicious code generation (Brundage et al. 2018).

Vulnerability discovery tools powered by AI further enhance cybercriminal capabilities. Machine learning systems can analyze software code to identify security weaknesses, enabling attackers to exploit vulnerabilities before they are patched. This significantly reduces the time window for defensive responses and increases system exposure. Adaptive malware represents a more advanced evolution of malicious software. Unlike static malware, adaptive systems can modify their behavior in response to detection attempts or system defenses. This dynamic capability makes traditional cybersecurity tools less effective, as malware can continuously evolve during execution.

E. Ransomware Evolution

Ransomware has also undergone significant transformation through the integration of GenAI technologies. Traditional ransomware attacks involved encryption of victim data followed by demands for payment. GenAI has introduced new layers of sophistication into this model.

Automated target selection allows ransomware systems to identify high-value targets based on data analysis. AI systems can evaluate organizational profiles, financial capacity, and vulnerability levels to prioritize potential victims. This increases the efficiency and profitability of ransomware campaigns.

AI negotiation systems represent a novel development in extortion strategies. In some cases, attackers deploy AI-driven chatbots to negotiate ransom payments with victims. These systems can adjust demands based on victim responses, financial capacity, and behavioral cues, thereby optimizing financial outcomes. Extortion optimization refers to the use of AI to maximize ransom payments through psychological manipulation, timing strategies, and adaptive communication. By analyzing victim behavior, AI systems can determine optimal pressure points to increase compliance rates (Europol 2022).

F. Information Manipulation and Disinformation

Generative AI has significantly expanded the scale and sophistication of information manipulation and disinformation campaigns. The ability to generate synthetic content at scale poses serious risks to democratic institutions, public discourse, and electoral integrity. Synthetic news generation enables the creation of entirely fabricated news articles that mimic legitimate journalistic outputs. These systems can produce realistic narratives that are difficult to distinguish from authentic reporting. Such capabilities can be exploited to spread misinformation or manipulate public opinion.

Political manipulation through GenAI involves the targeted dissemination of persuasive content designed to influence voter behavior, public sentiment, or policy debates. AI systems can tailor messages to specific demographic or ideological groups, increasing the effectiveness of

propaganda campaigns.

Election interference represents one of the most serious implications of GenAI-enabled disinformation. Deepfakes, automated bots, and synthetic media can be used to discredit candidates, distort political narratives, or suppress voter participation. These risks pose fundamental challenges to democratic legitimacy and electoral integrity (Chesney and Citron 2019).

G. Emergence of Autonomous Cybercrime Systems

A particularly significant and emerging development is the rise of semi-autonomous cybercrime systems. These systems represent a qualitative shift in the structure of cybercrime, moving beyond human-directed attacks toward AI-assisted or AI-orchestrated criminal operations. AI agents are increasingly capable of acting semi-independently, executing sequences of tasks without continuous human intervention. These agents can identify targets, generate attack strategies, and adapt tactics based on environmental feedback. This introduces the possibility of automated attack chains in which multiple stages of cybercrime are executed autonomously.

Automated attack chains refer to multi-step cyber operations conducted by interconnected AI systems. For example, one AI agent may identify vulnerabilities, another may generate exploit code, and a third may deploy phishing campaigns. This modular structure significantly increases operational efficiency and complexity. Self-improving attack systems represent the most advanced conceptual development in this domain. These systems are capable of learning from previous attacks, refining strategies, and improving success rates over time. While still largely theoretical, early research in autonomous AI agents suggests that such systems may become feasible in the near future (Russell and Norvig 2021; Brundage et al. 2018).

The emergence of autonomous cybercrime systems marks a critical turning point in cybersecurity governance. It challenges traditional assumptions regarding human intent, responsibility, and control in criminal law, and raises urgent questions about regulatory adequacy in the age of generative artificial intelligence.

LEGAL AND REGULATORY CHALLENGES

The rapid diffusion of Generative Artificial Intelligence (GenAI) in cybercrime ecosystems has exposed significant limitations in existing legal doctrines and regulatory architectures. Traditional criminal law frameworks were developed under assumptions of human-centered agency, linear causality, and territorially bounded jurisdiction. However, GenAI-enabled cybercrime disrupts these foundational assumptions by introducing distributed agency, automated decision-making, and borderless infrastructure. This section examines four core legal and regulatory challenges: attribution, mens rea, evidentiary reliability, jurisdictional complexity, and regulatory lag.

A. Attribution Problem in Criminal Law

One of the most fundamental challenges posed by GenAI-enabled cybercrime is the problem of attribution—identifying the legally responsible actor within complex socio-technical systems. Classical criminal law is premised on the principle that liability attaches to an identifiable human perpetrator who performs a voluntary act (*actus reus*) accompanied by a culpable mental state (*mens rea*). However, GenAI complicates this framework by introducing multi-layered causation chains involving users, developers, and platform providers.

At the most immediate level, users may deploy GenAI tools to generate malicious outputs such as phishing emails, deepfakes, or malware code. In such cases, users are often the most visible actors in the causal chain. However, their role may be limited to prompt engineering or indirect orchestration, raising questions about the extent of their intentional contribution to the harmful outcome.

Developers represent another potential locus of liability. AI model creators design architectures, training datasets, and system parameters that shape the behavior of generative systems. If harmful outputs are foreseeable consequences of model design, the question arises whether developers bear responsibility for negligent or reckless design choices. Nevertheless,

assigning criminal liability to developers is complicated by the general-purpose nature of most GenAI systems, which are designed for broad applicability rather than specific harmful uses (Cath 2018).

Platform providers constitute a third layer of potential responsibility. Cloud-based AI services and API providers facilitate access to generative models, often under conditions of limited control over downstream use. While platforms may implement safety filters or usage restrictions, their indirect involvement complicates the legal characterization of their role. The decentralization of responsibility across multiple actors creates a fragmented accountability structure. As a result, the causal chain in GenAI-enabled cybercrime is often diffuse, nonlinear, and probabilistic rather than direct and intentional. This undermines traditional doctrines of causation, complicating efforts to assign clear legal responsibility.

B. Mens Rea Problem

The doctrine of *mens rea*—the mental element of criminal liability—faces substantial challenges in the context of AI-assisted cybercrime. Traditional criminal law distinguishes between intentional, knowing, reckless, and negligent conduct. However, GenAI systems blur these distinctions by introducing varying degrees of automation and human control.

Intent in AI-assisted crime becomes difficult to establish when users rely on generative systems that produce outputs autonomously. A user may intend to generate benign content but inadvertently produce harmful material due to model unpredictability. Conversely, users may intentionally exploit GenAI capabilities for malicious purposes while maintaining plausible deniability regarding specific outputs.

Foreseeability versus automation further complicates the analysis. In legal doctrine, foreseeability is often used to determine whether a defendant should have anticipated the consequences of their actions. However, in GenAI systems, outputs are probabilistic and emergent, making precise prediction of outcomes difficult even for advanced users. This raises the question of whether users can reasonably foresee harmful outputs generated by complex models (Wachter, Mittelstadt, and Floridi 2017).

Recklessness in digital environments becomes particularly relevant in this context. Recklessness traditionally involves conscious disregard of a substantial and unjustifiable risk. In GenAI systems, however, risk assessment is mediated by technological opacity and uncertainty. Users may lack sufficient understanding of model behavior to accurately assess risks, thereby challenging the applicability of conventional recklessness standards. The result is a doctrinal tension between subjective mental states and objective technological complexity. Criminal law struggles to determine whether liability should be based on intent, knowledge, or mere usage of dual-use technologies.

C. Evidentiary Challenges

GenAI introduces significant evidentiary challenges for criminal justice systems, particularly in relation to the authenticity, reliability, and interpretability of digital evidence. One major challenge is the detection of AI-generated content. Deepfake technologies and advanced language models can produce highly realistic text, audio, and visual materials that are indistinguishable from authentic human-generated content. Traditional forensic methods, which rely on metadata analysis or pattern recognition, are increasingly insufficient to reliably detect synthetic content (Farid 2022).

Authenticity verification becomes a central concern in legal proceedings. Courts must determine whether digital evidence reflects genuine events or artificially generated artifacts. This requires advanced forensic tools capable of tracing provenance, analyzing generation patterns, and verifying cryptographic integrity. However, such tools are not yet universally available or standardized across jurisdictions.

Digital forensics limitations further exacerbate these challenges. The rapid evolution of GenAI systems outpaces the development of forensic methodologies, creating a technological asymmetry between offense and defense capabilities. Additionally, encrypted communications, decentralized storage systems, and cloud-based infrastructures complicate the retrieval and preservation of evidence.

These evidentiary challenges have direct implications for due process

rights, as unreliable or unverifiable evidence may lead to wrongful convictions or procedural unfairness. The burden of proof becomes increasingly difficult to satisfy in environments where the distinction between real and synthetic data is blurred.

D. Jurisdictional Complexity

GenAI-enabled cybercrime operates within a fundamentally borderless digital infrastructure, creating complex jurisdictional challenges for enforcement and adjudication. Cross-border infrastructure is a defining feature of contemporary cybercrime ecosystems. Attacks may originate in one jurisdiction, be executed through servers in another, and impact victims located in multiple countries simultaneously. This fragmentation complicates the application of territorial jurisdiction principles traditionally used in criminal law.

Cloud computing further exacerbates jurisdictional ambiguity. GenAI systems are often hosted on distributed cloud infrastructures that dynamically allocate computational resources across multiple regions. As a result, the physical location of data processing becomes fluid and difficult to determine. Server location ambiguity presents additional challenges. Criminal investigations often rely on identifying the physical location of servers involved in cyberattacks. However, virtualization technologies and distributed networks obscure server locations, making it difficult to establish jurisdictional authority over relevant infrastructure.

These factors collectively undermine traditional jurisdictional doctrines based on territoriality and sovereignty. They necessitate increased reliance on international cooperation mechanisms, mutual legal assistance treaties, and transnational regulatory frameworks (Clough 2015).

E. Regulatory Lag

A central structural challenge in governing GenAI-enabled cybercrime is regulatory lag—the persistent delay between technological innovation and legal adaptation. This phenomenon reflects the inherent asymmetry between rapid technological development and the slower processes of legislative

reform.

Historically, law has been reactive rather than anticipatory, responding to technological developments only after their societal impacts become apparent. This reactive model is increasingly inadequate in the context of GenAI, where technological evolution occurs at exponential speed and with unpredictable trajectories (Lessig 2006).

The distinction between reactive law and anticipatory governance is therefore critical. Reactive legal frameworks attempt to regulate harms after they occur, while anticipatory governance seeks to identify and mitigate risks before they materialize. In the context of GenAI, anticipatory governance is increasingly necessary to address emergent threats such as autonomous cybercrime systems and large-scale disinformation campaigns.

Regulatory lag is further exacerbated by the global nature of AI development. Competing regulatory regimes, differing national priorities, and rapid private-sector innovation contribute to fragmented governance landscapes. As a result, legal systems often struggle to keep pace with technological change, creating gaps in enforcement and accountability. Ultimately, regulatory lag underscores the need for adaptive, flexible, and forward-looking legal frameworks capable of responding to continuous technological transformation rather than static regulatory models.

ASSESSMENT OF EXISTING LEGAL FRAMEWORKS

The rapid expansion of Generative Artificial Intelligence (GenAI) has exposed significant limitations in existing cybercrime governance frameworks across international, regional, and national levels. While various jurisdictions have developed regulatory responses to digital crime and artificial intelligence, most existing instruments were designed prior to the emergence of GenAI and therefore remain inadequately equipped to address its transformative impact. This section critically evaluates major legal frameworks, including international conventions, the European Union's regulatory model, the United States' sectoral approach, China's centralized governance system, and ASEAN–Indonesia's emerging legal architecture.

A. International Frameworks: Budapest Convention on Cybercrime

The Budapest Convention on Cybercrime (2001) remains the most influential international treaty addressing cybercrime. It establishes a harmonized legal framework for criminalizing offenses such as illegal access, data interference, system interference, and computer-related fraud, while also promoting international cooperation in investigation and prosecution (Council of Europe 2001).

1) Strengths

   The primary strength of the Budapest Convention lies in its role as a foundational instrument for harmonizing cybercrime legislation across jurisdictions. It provides standardized definitions of core cyber offenses and facilitates cross-border cooperation through mechanisms such as mutual legal assistance and expedited data sharing. This has significantly improved the ability of states to respond to transnational cyber threats. Another important strength is its procedural flexibility. The Convention does not prescribe rigid technological requirements, allowing states to adapt its provisions to evolving digital environments. This adaptability has contributed to its widespread adoption and continued relevance in conventional cybercrime cases.

2) Limitations for AI-Enabled Cybercrime

   Despite its strengths, the Budapest Convention was designed in an era preceding the rise of artificial intelligence. As such, it does not account for autonomous or semi-autonomous systems capable of generating harmful content. The Convention assumes human intentionality and direct action, which are increasingly absent in GenAI-mediated cybercrime. Furthermore, the treaty lacks provisions addressing algorithmic accountability, AI-generated evidence, or machine-driven decision-making processes. It also does not address the problem of synthetic content creation, such as deepfakes or AI-generated phishing materials. As a result, its applicability to GenAI-enabled cybercrime is limited and increasingly outdated in practice (Schjølberg 2017).

B. European Union Approach

The European Union has developed one of the most comprehensive and rights-based regulatory frameworks for artificial intelligence and digital governance. Its approach is characterized by a strong emphasis on risk regulation, fundamental rights protection, and ex ante governance mechanisms.

1) AI Act

   The EU Artificial Intelligence Act represents a landmark regulatory instrument that introduces a risk-based classification system for AI technologies. High-risk AI systems, including those used in law enforcement and critical infrastructure, are subject to strict obligations regarding transparency, documentation, human oversight, and conformity assessments (Veale and Borgesius 2021). This framework is particularly relevant for GenAI-enabled cybercrime because it explicitly addresses systemic risks associated with advanced AI models. However, its primary focus remains on regulatory compliance rather than criminal liability, limiting its direct applicability to cybercrime enforcement.

2) Digital Services Act

   The Digital Services Act (DSA) complements the AI Act by regulating online platforms and intermediaries. It imposes obligations on large platforms to assess systemic risks, remove illegal content, and ensure transparency in content moderation systems. The DSA also introduces accountability mechanisms for algorithmic recommendation systems. In the context of cybercrime, the DSA plays a critical role in addressing the dissemination of illegal content, including synthetic media and disinformation. However, it primarily targets platform governance rather than criminal conduct itself.

3) Risk-Based Regulation Model

   The EU's regulatory architecture is grounded in a risk-based model that differentiates between unacceptable, high, limited, and minimal risk AI systems. This approach enables proportional regulation based on potential harm to fundamental rights and societal interests. It represents a shift from reactive enforcement to preventive governance, emphasizing ex

ante assessment and compliance obligations.

C. United States Approach

The United States adopts a fragmented and sector-specific approach to cybercrime and artificial intelligence regulation. Unlike the EU's unified legislative model, the U.S. relies on a combination of federal statutes, state laws, and executive actions.

1) Sectoral Regulation
   Cybercrime regulation in the United States is distributed across multiple legal instruments, including the Computer Fraud and Abuse Act (CFAA), state-level privacy laws, and sector-specific regulations in finance, healthcare, and communications. While this approach allows flexibility, it also creates inconsistencies and regulatory gaps. In relation to AI, there is no comprehensive federal law governing artificial intelligence systems. Instead, regulation is dispersed across agencies such as the Federal Trade Commission (FTC), which enforces consumer protection laws against deceptive AI practices.
2) Executive Orders on AI Safety
   Recent executive orders have introduced policy guidelines for AI safety, security, and ethical development. These directives emphasize risk management, transparency, and testing requirements for advanced AI systems developed by private companies. However, executive orders lack the permanence and enforceability of statutory law, limiting their long-term regulatory impact.
3) Fragmented Cyber Law System
   The U.S. cyber legal framework is characterized by fragmentation and overlapping jurisdictional authority. This creates challenges in addressing GenAI-enabled cybercrime, particularly in ensuring consistent standards for accountability, detection, and enforcement across different sectors and states (Citron 2022).

D. China Approach

China represents a highly centralized and state-controlled model of AI

governance. Its regulatory approach emphasizes security, social stability, and strict control over digital platforms.

1) Generative AI Regulation
   China has introduced specific regulations governing generative AI services, requiring providers to ensure that outputs are accurate, lawful, and aligned with state values. These regulations impose strict licensing requirements and content moderation obligations on AI developers and service providers.
2) Content Labeling Rules
   A key feature of China's regulatory framework is mandatory content labeling for AI-generated material. This includes requirements to clearly identify synthetic content, such as deepfakes and AI-generated text or images. This measure aims to reduce misinformation and enhance transparency in digital environments.
3) Strict Platform Control
   Chinese authorities maintain strong regulatory oversight over digital platforms, including real-time monitoring and enforcement mechanisms. Platforms are required to implement robust content filtering systems and cooperate with government agencies in detecting and removing harmful content. While effective in enforcement, this approach raises concerns regarding freedom of expression and privacy.

E. ASEAN and Indonesia

In Southeast Asia, including Indonesia, regulatory frameworks for cybercrime and artificial intelligence remain in developmental stages. Existing laws primarily focus on general cybercrime and data protection rather than AI-specific governance.

1) UU ITE
   Indonesia's Electronic Information and Transactions Law (UU ITE) provides the primary legal basis for cybercrime regulation. It criminalizes unauthorized access, online defamation, and electronic fraud. However, it does not explicitly address AI-generated content or autonomous cybercrime systems.

2) UU PDP
   The Personal Data Protection Law (UU PDP) establishes important safeguards for data privacy, including principles of lawful processing, data minimization, and data subject rights. While relevant to AI governance, it does not directly regulate the use of generative models or algorithmic decision-making in cybercrime contexts.
3) KUHP Baru
   The new Indonesian Criminal Code (*KUHP Baru*) modernizes several aspects of criminal law but remains largely technology-neutral. It does not contain specific provisions addressing AI-enabled cybercrime, deepfakes, or automated criminal systems.

Overall, ASEAN-level cooperation in cybercrime governance remains limited, with no binding regional framework comparable to the Budapest Convention or EU regulatory architecture.

F. Key Regulatory Gap

Despite the diversity of legal frameworks across jurisdictions, several critical regulatory gaps remain evident in relation to GenAI-enabled cybercrime. First, there is a general absence of AI-specific cybercrime provisions in most legal systems. Existing laws focus on human-perpetrated digital offenses and do not adequately address autonomous or semi-autonomous AI-generated criminal activity. Second, there is no universal obligation for AI system developers or operators to implement detection mechanisms for malicious use. This creates enforcement blind spots, particularly in identifying AI-generated phishing, deepfakes, and automated attack systems.

Third, algorithmic accountability rules remain underdeveloped in most jurisdictions. There is limited legal requirement for transparency, auditability, or explainability of AI systems used in cybercrime contexts. This weakens both preventive and corrective regulatory mechanisms. Collectively, these gaps demonstrate that existing legal frameworks are not yet fully equipped to address the structural transformation of cybercrime driven by

Generative Artificial Intelligence. This reinforces the need for a more integrated, adaptive, and risk-based global regulatory approach.

TOWARD A MULTI-LAYERED REGULATORY FRAMEWORK

The rapid evolution of Generative Artificial Intelligence (GenAI) and its integration into cybercrime ecosystems necessitate a fundamental rethinking of existing regulatory architectures. Traditional single-layer regulatory models, which rely predominantly on criminal law enforcement or sectoral regulation, are increasingly inadequate in addressing the distributed, adaptive, and transnational nature of AI-enabled cyber threats. Consequently, a multi-layered regulatory framework is required to ensure that governance mechanisms operate across different levels of the digital ecosystem, ranging from individual criminal liability to global institutional cooperation. This section develops such a framework by integrating normative principles with four interconnected layers of regulation: criminal law reform, platform accountability, AI system governance, and international cooperation.

A. Normative Principles

The foundation of any effective regulatory response to GenAI-enabled cybercrime must be grounded in a set of normative principles that guide both legislative design and institutional implementation. First, the principle of technology neutrality requires that legal frameworks avoid over-specification of particular technologies, ensuring that regulations remain applicable regardless of technological evolution. This principle is essential in the context of rapidly changing AI systems, where specific technical definitions may quickly become obsolete (Lessig 2006).

Second, proportionality serves as a balancing mechanism between regulatory intervention and technological innovation. Regulatory measures must be proportionate to the severity and likelihood of harm posed by GenAI-enabled cybercrime. Overly restrictive frameworks may stifle innovation, while under-regulation may expose societies to significant cyber risks.

Proportionality therefore ensures that regulatory responses are calibrated to actual risk levels rather than hypothetical fears.

Third, risk-based governance provides a structured approach to regulation by categorizing AI systems and cyber threats according to their potential impact on individuals, institutions, and society. High-risk applications, such as AI-driven identity fraud or automated cyberattacks, require stricter oversight mechanisms compared to low-risk applications. This approach aligns with emerging global regulatory trends, particularly in the European Union's AI governance model (Veale and Borgesius 2021).

Fourth, human rights protection constitutes the normative core of the proposed framework. Any regulatory response to GenAI-enabled cybercrime must be consistent with fundamental rights, including privacy, due process, freedom of expression, and non-discrimination. This ensures that cybersecurity measures do not themselves become sources of rights violations, particularly in contexts involving surveillance, automated profiling, or content moderation.

B. Layer 1: Criminal Law Reform

The first layer of the proposed framework focuses on reforming substantive criminal law to address the unique characteristics of AI-enabled cybercrime. Existing criminal statutes are primarily designed to address human-perpetrated offenses and therefore require adaptation to capture emerging forms of technologically mediated criminality.

One important reform is the introduction of aggravating circumstances for the use of AI in the commission of cybercrime. When artificial intelligence is used to enhance the scale, sophistication, or impact of criminal conduct, sentencing frameworks should reflect the increased harm potential. This approach recognizes that AI does not merely facilitate crime but can significantly amplify its effects, particularly in cases involving automated phishing campaigns, deepfake fraud, or large-scale disinformation operations.

Another critical reform is the criminalization of synthetic fraud. Synthetic fraud refers to the creation and use of AI-generated identities, documents, or

audiovisual materials for deceptive purposes. Unlike traditional fraud, which relies on the misrepresentation of existing identities, synthetic fraud involves the fabrication of entirely new digital personas. This phenomenon poses unique challenges for legal classification and requires explicit statutory recognition to ensure effective enforcement.

In addition, the legal system must recognize AI-enabled cybercrime offenses as a distinct category of criminal conduct. This includes offenses where AI systems are used to automate hacking, generate malware, execute phishing campaigns, or manipulate digital environments. Such recognition would provide clearer legal grounds for prosecution and reduce interpretive uncertainty in judicial proceedings. Overall, criminal law reform constitutes the foundational enforcement layer, ensuring that liability frameworks remain relevant in the face of technological transformation.

C. Layer 2: Platform Accountability

The second layer of the framework addresses the role of digital platforms, cloud service providers, and AI system intermediaries in facilitating or mitigating cybercrime. Given that GenAI systems are often accessed through platform-based infrastructures, these entities occupy a critical position in the cybercrime ecosystem.

A central component of platform accountability is the establishment of detection obligations. Platforms should be legally required to implement mechanisms for identifying AI-generated malicious content, including phishing messages, deepfake media, and automated scam operations. These detection systems must be continuously updated to reflect evolving threat patterns and adversarial techniques.

Reporting mechanisms constitute another essential element of this layer. Platforms should be obligated to report significant cyber threats, breaches, or malicious AI usage to relevant authorities. Such mechanisms enable early intervention and enhance collective cybersecurity resilience by facilitating information sharing between private and public actors.

Transparency requirements further strengthen platform accountability by mandating disclosure of AI system functionalities, content moderation

policies, and risk management practices. Transparency ensures that users, regulators, and oversight bodies can assess the reliability and safety of AI systems. In the absence of transparency, accountability becomes difficult to enforce, particularly in cases involving proprietary or black-box systems. Collectively, these obligations transform platforms from passive intermediaries into active participants in cybersecurity governance.

D. Layer 3: AI Governance

The third layer focuses on the governance of AI systems themselves, particularly during their design, development, and deployment phases. This preventive layer is essential for addressing risks before they materialize into criminal harm. Safety-by-design principles require that AI systems be developed with embedded safeguards to prevent misuse. This includes restricting harmful outputs, implementing content filters, and designing systems that are resistant to adversarial manipulation. Safety-by-design shifts responsibility upstream to the development stage, where risks can be mitigated more effectively.

Security-by-design complements this approach by ensuring that AI systems are resilient against external attacks, including data poisoning, model inversion, and prompt injection. In the context of GenAI, security vulnerabilities can be exploited to generate malicious outputs or bypass safety mechanisms, making robust security architecture essential. Red teaming represents a critical operational mechanism for evaluating AI system vulnerabilities. This process involves simulating adversarial attacks to identify weaknesses in model behavior, safety constraints, and output reliability. Regular red teaming exercises help ensure that AI systems remain robust against evolving cyber threats.

Model auditing provides an additional layer of oversight by enabling independent evaluation of AI systems' performance, bias, and compliance with regulatory standards. Audits should assess not only technical accuracy but also ethical and legal implications, particularly in high-risk applications involving cybersecurity or criminal justice. Together, these mechanisms establish a preventive governance structure that reduces the likelihood of AI

systems being exploited for cybercriminal purposes.

E. Layer 4: International Cooperation

Given the inherently transnational nature of cybercrime, international cooperation constitutes the fourth and indispensable layer of the regulatory framework. GenAI-enabled cybercrime often involves distributed infrastructures, cross-border data flows, and actors operating in multiple jurisdictions, making unilateral enforcement insufficient.

Mutual legal assistance mechanisms remain central to international cooperation. These mechanisms facilitate cross-border evidence sharing, extradition, and joint investigations. However, existing frameworks must be modernized to accommodate the speed and complexity of AI-enabled cybercrime investigations. Cross-border cybercrime coordination should be strengthened through real-time information sharing platforms and joint response teams. Such coordination enables rapid identification and mitigation of large-scale cyber threats, particularly those involving automated or distributed attack systems.

Global AI governance standards are also necessary to harmonize regulatory approaches across jurisdictions. Without common standards, regulatory fragmentation may create enforcement gaps that can be exploited by cybercriminal actors. International organizations and multilateral institutions play a crucial role in developing these shared norms.

F. Core Contribution: Multi-Layer Governance Model

The central contribution of this article is the formulation of a multi-layer governance model designed to address the structural complexity of GenAI-enabled cybercrime. This model integrates four interconnected regulatory layers that operate simultaneously across different levels of the digital ecosystem.

The first layer is criminal law enforcement, which establishes substantive liability for AI-enabled cybercrime and defines aggravating circumstances for technological facilitation. The second layer is platform regulation, which imposes obligations on intermediaries to detect, report, and mitigate cyber

threats. The third layer is AI system governance, which ensures that AI technologies are developed and deployed in a safe, secure, and auditable manner. The fourth layer is international cooperation, which enables cross-border coordination and harmonization of regulatory standards. Together, these four layers form a comprehensive governance architecture that addresses both the causes and consequences of GenAI-enabled cybercrime. By integrating preventive, operational, and reactive mechanisms, the multi-layered framework provides a more resilient and adaptive regulatory response to the evolving challenges of artificial intelligence in cyberspace.

FUTURE DIRECTIONS OF GENAI CYBERCRIME

The trajectory of Generative Artificial Intelligence (GenAI) indicates that cybercrime will continue to evolve beyond its current forms of automation and augmentation toward increasingly autonomous, adaptive, and self-sustaining systems. While present manifestations of GenAI-enabled cybercrime still largely depend on human initiation and oversight, emerging technological developments suggest a future in which artificial intelligence systems may assume more active, coordinated, and potentially semi-independent roles in cybercriminal ecosystems. This section explores four interrelated future directions: autonomous AI agents in cybercrime, agentic AI systems, the emergence of self-evolving cybercrime ecosystems, and the need for anticipatory regulation.

Autonomous AI agents represent a significant shift in the architecture of cyber operations. Unlike conventional generative models that respond to discrete user prompts, autonomous agents are capable of executing multi-step tasks, maintaining memory, and pursuing predefined objectives with minimal human intervention. In cybercrime contexts, such agents could theoretically be deployed to automate the entire lifecycle of an attack, including reconnaissance, vulnerability identification, exploitation, and post-exploitation activities.

The primary concern is that autonomous agents reduce the need for continuous human supervision. A single operator may initiate a malicious

objective, after which AI agents independently perform complex sequences of actions across digital environments. This includes scanning networks, identifying weak points, generating exploit code, and deploying phishing campaigns. The result is a shift from manual cybercrime operations to delegated cybercrime execution systems, where human involvement becomes increasingly abstract and remote (Russell and Norvig 2021; Brundage et al. 2018). Such developments raise profound legal and ethical questions regarding responsibility and control. If an autonomous agent executes harmful actions beyond the immediate intent or control of its human operator, traditional doctrines of criminal liability become difficult to apply.

Closely related to autonomous agents is the concept of agentic AI systems, which refers to AI models designed not only to generate outputs but also to make decisions, plan strategies, and adapt dynamically to environmental feedback. These systems are characterized by goal-oriented behavior, recursive reasoning, and the ability to interact with other digital systems in coordinated ways. In cybercrime scenarios, agentic AI systems could function as strategic coordinators that manage multiple sub-agents performing specialized tasks. For example, one agent may focus on target selection, another on vulnerability exploitation, and another on data exfiltration or extortion strategies. This modular architecture increases efficiency and resilience, as the failure of one component does not necessarily disrupt the entire operation.

Agentic systems also introduce adaptive learning capabilities, allowing cybercriminal operations to evolve based on defensive responses. If a particular attack vector is blocked, the system may automatically adjust its strategy and attempt alternative methods. This adaptive behavior significantly increases the difficulty of detection and mitigation by traditional cybersecurity systems, which are often designed to respond to static threat patterns.

From a legal perspective, agentic AI systems blur the boundary between tool and actor. As these systems become more sophisticated, they challenge existing assumptions that AI is merely an instrument of human will rather than an independent or semi-independent participant in cyber activities.

One of the most concerning long-term scenarios is the emergence of self-evolving cybercrime ecosystems. These ecosystems refer to interconnected networks of AI agents, tools, and infrastructures that continuously learn, adapt, and optimize their criminal strategies without centralized human control. In such ecosystems, different AI components may interact in a decentralized manner, sharing data, refining attack techniques, and iteratively improving performance. For instance, one AI system may generate phishing templates, another may analyze their success rates, and a third may optimize future campaigns based on feedback loops. Over time, this creates a form of evolutionary optimization within cybercriminal operations.

The risk of such systems lies in their potential autonomy and scalability. Once established, self-evolving ecosystems may operate with minimal human oversight, making them difficult to dismantle or regulate. Moreover, their adaptive nature allows them to respond dynamically to defensive measures, creating a persistent and evolving threat landscape. This scenario represents a qualitative transformation in cybercrime structure, shifting from human-directed criminal enterprises to algorithmically driven ecosystems. It also raises significant concerns regarding unpredictability, as emergent behaviors may arise that were not explicitly programmed or anticipated by developers or operators (Chesney and Citron 2019; Brundage et al. 2018).

The emergence of autonomous agents, agentic systems, and self-evolving cybercrime ecosystems underscores the urgent need for anticipatory regulation. Traditional reactive legal frameworks, which intervene only after harm has occurred, are increasingly insufficient in addressing the speed and complexity of GenAI-driven threats. Anticipatory regulation emphasizes proactive governance strategies that seek to identify, assess, and mitigate risks before they materialize into large-scale harm. This includes the development of early warning systems, scenario-based risk assessments, and continuous monitoring of AI capabilities as they evolve.

In the context of cybercrime, anticipatory regulation may involve requiring developers and deployers of advanced AI systems to conduct rigorous risk evaluations prior to release. It may also include mandatory safety testing for autonomous agent architectures, particularly those capable

of executing multi-step actions without human intervention.

Furthermore, anticipatory governance requires international coordination, as cyber threats are inherently transnational. Regulatory frameworks must evolve in parallel across jurisdictions to prevent regulatory arbitrage, where malicious actors exploit weaker legal environments. Ultimately, anticipatory regulation represents a paradigm shift from reactive enforcement to preventive governance. It recognizes that in the era of GenAI, waiting for harm to occur before responding is no longer a viable strategy for ensuring cybersecurity, legal accountability, and public safety.

CONCLUSION

The analysis demonstrates that Generative Artificial Intelligence fundamentally transforms the structure of cybercrime rather than merely enhancing existing criminal techniques. GenAI significantly increases the scale, speed, and sophistication of cybercriminal activities by enabling automation, hyper-personalization, and real-time adaptability across multiple offense types, including phishing, identity fraud, financial scams, malware generation, ransomware optimization, and disinformation campaigns. These developments collectively indicate that cybercrime is no longer predominantly human-centered but increasingly AI-augmented and partially automated. At the same time, the study shows that existing legal frameworks—both at international and national levels—remain insufficient to address these transformations, as they are still grounded in assumptions of human intent, linear causality, and territorially bounded enforcement. This mismatch generates substantial gaps in attribution, evidentiary standards, jurisdictional authority, and regulatory capacity, thereby weakening the effectiveness of traditional criminal justice responses to emerging digital threats.

Building on these findings, the article develops the theoretical contribution of "AI-amplified cybercrime," which conceptualizes cybercrime as a phenomenon structurally enhanced by artificial intelligence in terms of operational capacity and systemic impact. This concept captures the shift

from technology-assisted crime to AI-driven crime ecosystems characterized by automation and adaptive learning dynamics. In addition, the article advances a policy contribution in the form of a multi-layer regulatory framework that integrates criminal law reform, platform accountability, AI system governance, and international cooperation. The final argument is that regulatory approaches must move beyond reactive criminal law enforcement toward anticipatory, risk-based, and multi-layer governance models capable of addressing the evolving nature of AI-enabled cybercrime. Such a shift is essential to ensure that legal systems remain effective, resilient, and aligned with the technological realities of the GenAI era.

ACKNOWLEDGMENTS

None.

FUNDING INFORMATION

None.

CONFLICTING INTEREST AND ORGINALITY STATEMENT

The authors state that there is no conflict of interest in the publication of this article.

GENERATIVE AI STATEMENT

None.

REFERENCES

Bishop, Matt. 2003. *Computer Security: Art and Science*. Boston: Addison-Wesley.

Brenner, Susan W. 2010. *Cybercrime: Criminal Threats from Cyberspace*. Santa Barbara: Praeger.

Brown, Tom B., Benjamin Mann, Nick Ryder, et al. 2020. "Language Models are Few-Shot Learners." In *Advances in Neural Information Processing*

Systems. (Book/Proceedings volume)

Brundage, Miles, Shahar Avin, Jack Clark, et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Oxford: Future of Humanity Institute.

Cath, Corinne. 2018. "Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges." *Philosophical Transactions of the Royal Society A* 376 (2133).

Chen, Mark, Jerry Tworek, Heewoo Jun, et al. 2021. *Evaluating Large Language Models Trained on Code*. OpenAI Technical Report.

Chesney, Robert, and Danielle Keats Citron. 2019. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107 (6): 1753–1820.

Citron, Danielle Keats. 2022. *The Fight for Privacy: Protecting Dignity, Identity, and Trust in the Digital Age*. New York: Norton.

Clarke, Richard A., and Robert Knake. 2010. *Cyber War*. New York: HarperCollins.

Clough, Jonathan. 2015. *Principles of Cybercrime*. Cambridge: Cambridge University Press.

Council of Europe. 2001. *Convention on Cybercrime (Budapest Convention)*.

European Union. 2022. *Artificial Intelligence Act (Draft Regulation)*.

European Union. 2022. *Digital Services Act (Regulation (EU) 2022/2065)*.

Europol. 2022. *Internet Organised Crime Threat Assessment (IOCTA) 2022*. The Hague: Europol.

Farid, Hany. 2022. *Fake Photos and Real Liars: The Rise of Deepfakes*. MIT Press.

Floridi, Luciano, and Massimo Chiriatti. 2020. "GPT-3: Its Nature, Scope, Limits, and Consequences." *Minds and Machines* 30 (4): 681–694.

Floridi, Luciano. 2014. *The Ethics of Information*. Oxford: Oxford University Press.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge: MIT Press.

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. "Denoising Diffusion Probabilistic Models." In *NeurIPS Proceedings*.

Jia, Ye, et al. 2018. *Transfer Learning from Speaker Verification to Multispeaker*

Text-To-Speech Synthesis. Google Research Technical Report.
Kshetri, Nir. 2013. *Cybercrime and Cybersecurity in the Global South*. Springer.
Lessig, Lawrence. 2006. *Code: Version 2.0*. New York: Basic Books.
Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
OECD. 2019. *OECD Principles on Artificial Intelligence*. Paris: OECD Publishing.
Ohm, Paul. 2010. "Broken Promises of Privacy." *UCLA Law Review* 57: 1701–1777.
Russell, Stuart, and Peter Norvig. 2021. *Artificial Intelligence: A Modern Approach*. 4th ed. Harlow: Pearson.
Schjølberg, Stein. 2017. *Cybercrime Law and AI Governance*. Oslo: Norwegian Center for Cybersecurity Law.
State Council of China. 2023. *Interim Measures for the Management of Generative Artificial Intelligence Services*.
U.S. Congress. 1986. *Computer Fraud and Abuse Act (CFAA)*.
UNESCO. 2021. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.
Veale, Michael, and Frederik Zuiderveen Borgesius. 2021. "Demystifying the Draft EU AI Act." *Computer Law Review International* 22 (4): 97–112.
Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR." *International Data Privacy Law* 7 (2): 76–99.
Wall, David S. 2007. *Cybercrime: The Transformation of Crime in the Information Age*. Cambridge: Polity Press.
White House. 2023. *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*.
Yar, Majid. 2013. *Cybercrime and Society*. London: Sage.
Zhang, Chen, and Dawei Li. 2022. *Internet Governance and Cybersecurity in China*. Beijing: Springer.

This page intetionally left blank