From Phishing to Deepfake: The Rapid Evolution of Cybercrime Tactics

Watno Saputro[1], Hani Irhamdessetya[2]
Bachelor Of Law, Universitas Ngudi Waluyo, Indonesia[1],
Master Of Law, Universitas Ngudi Waluyo, Indonesia[2]
Email Correspondence: haniirhamdessetya@unw.ac.id[2]

Abstract

This study analyzes the fundamental shift in cybercrime tactics, moving from conventional Phishing, which exploits human vulnerabilities, toward sophisticated Artificial Intelligence (AI)-based attacks like Deepfake. Phishing, though simple, remains a dominant attack vector ; however, Deepfake, powered by Generative Adversarial Networks (GANs), creates false "digital evidence" with near-perfect credibility. This transition escalates the risk of catastrophic losses, threatening the integrity of digital evidence and institutional stability. Employing the Normative Legal Research method, this study evaluates the defense gap. Current cybersecurity strategies, focused on malware mitigation and user education, are proven ineffective against AI-based threats. The conclusion emphasizes the urgent need for a paradigm shift toward proactive, integrity-based defense and recommends revising criminal and civil procedural law for digital evidence authentication, alongside clear regulations on Deepfake accountability

Keywords: Cybercrime; Phishing; Deepfake; Artificial Intelligence (AI); Digital Evidence.

Introduction

The modern digital world, which forms the backbone of the global economy and social interaction, is now facing a security crisis whose escalation is increasingly uncontrolled. Cybercrime has gone beyond the boundaries of mere technical threats; it has transformed(Wall, 2024) into a highly organized global dark economy, with projected losses reaching trillions of dollars in the next few years. For a long time, the foundation of almost all cybercrime has been Phishing. This tactic, which relies on social engineering and psychological manipulation, is relatively simple yet offers a very high Return on Investment (ROI) for its perpetrators(Cuganesan & Lacey, 2011). Phishing exploits the most fundamental vulnerability in the security chain the human factor through fake emails, text messages, or phone calls designed to steal login credentials and sensitive information. Although considered "classic," Phishing remains the dominant initial (Alkhalil et al., 2021) attack vector to this day.

However, the cyber threat ecosystem is never static. As companies increase their investment in technical security (such as firewalls and antivirus software), criminals respond with a rapid and sophisticated metamorphosis of tactics. The next evolutionary phase brought an increase in more targeted attacks, known as Spear Phishing and Whaling, where messages are customized specifically for high-profile victims such as C-level executives. Furthermore, the emergence of a new generation of Ransomware marked a turning point. Perpetrators no longer just steal data but hold entire digital infrastructures hostage often with a double extortion model (data encryption and the threat of leaking the stolen data). This development is supported by the rise of Cybercrime as a Service (CaaS) on the Dark Web(Akyazi et al., 2021), which democratizes access to advanced hacking tools, enabling almost anyone with a small capital to launch complex attacks.

The most revolutionary transition in cybercrime tactics occurred with the integration of Artificial Intelligence (AI), specifically in the form of Deepfake(Westerlund, 2019). Deepfake represents a quantum leap from text/link-based Phishing to highly realistic multimedia-based manipulation(Westerlund, 2019). If Phishing relies on a trick that can be identified through careful scrutiny, Deepfake creates a nearly perfect alternative reality. With AI’s ability to generate incredibly convincing fake videos, audio, and images, Deepfake threatens to collapse the boundaries between digital truth and falsehood.

This poses serious consequences:

a. Criminals can replicate a CEO's voice (voice cloning) to order an urgent fund transfer to a junior employee, resulting in instantaneous and significant financial losses.
b. Deepfake can be used to create credible fake videos, damaging individual reputations, or manipulating stock markets or political processes, threatening institutional stability.

c. Security systems that use facial or voice recognition can be compromised by highly accurate Deepfakes.

This threat goes far beyond the scope of Phishing. It is a threat that fundamentally damages the integrity of digital evidence and public trust.

Although Phishing has been extensively studied and Deepfake is beginning to receive increased attention, the literature still lacks a comprehensive and structured analysis of the evolutionary trajectory of cybercrime itself, connecting conventional Phishing with AI-based Deepfake threats. Many studies tend to focus on one specific tactic rather than understanding the continuity of increasing sophistication and its impact.

The sharp transition from Phishing to Deepfake clearly indicates that cybercrime has entered a new era, where technological sophistication is the primary determinant of attack effectiveness.(Collier & Clayton, 2022) While previous defenses could focus on user education and link filtering, the threat now stems from the manipulation of digital reality itself, demanding a fundamentally different response. This shift creates a critical gap between rapidly evolving attack methods and the readiness of current security systems.

Therefore, in order to fully comprehend the scale of this threat and design adaptive defense strategies, fundamental questions emerge that must be addressed structurally.

Based on the preceding background, this study formulates the main research problems as follows: How has the evolutionary trajectory of cybercrime tactics shifted from conventional social engineering methods based on human vulnerabilities (such as Phishing) toward highly sophisticated Artificial Intelligence (AI)-based attacks capable of multimedia manipulation (such as Deepfake), and what are the fundamental differences in complexity, credibility, and potential loss between these two generations of attacks? And To what extent are current cybersecurity strategies and technologies, predominantly focused on mitigating conventional threats (Phishing and malware), effective in addressing Deepfake and other emerging AI-based cyber threats, and what proactive adaptations are necessary within the digital security and identity verification frameworks to maintain the integrity of digital evidence and public trust?

Research Method

This study employs the Normative Legal Research (Yuridis Normatif) method with a qualitative approach. Normative legal research is chosen because the article's main focus is analyzing the gap between the evolving cybercrime phenomena (from Phishing to Deepfake) and the applicable legal framework, policies, and mitigation strategies.

The specific approaches utilized are:

a) Conceptual Approach: To analyze core concepts related to the evolution of cybercrime tactics, social engineering, and the implications of Artificial Intelligence (AI) for digital security.
b) Statute Approach: To identify and review relevant national and international legislation concerning data protection, cybercrime (e.g., the ITE Law), and digital authentication regulations, in order to assess the legal sufficiency in responding to the **Deepfake** threat.

The type of data used is secondary data, which encompasses primary legal materials (statutes and regulations), secondary legal materials (scholarly journals, expert publications, and books related to cybersecurity and law), and tertiary legal materials (dictionaries and encyclopedias). The data are qualitatively analyzed using systematic interpretation techniques and deductive logic to draw conclusions regarding the legal implications and adaptive policy recommendations against new-generation cyber threats.

Disucussion

Analysis of the Evolutionary Trajectory of Cybercrime Tactics

This problem focuses on analyzing the fundamental shift in the cyber threat landscape, from methods relying on human psychological manipulation to attacks driven by technological sophistication. The evolution of these cybercrime tactics can be understood through an in-depth comparison between Phishing (the conventional generation) and Deepfake (Schmitt & Flechais, 2024)(the AI-based generation), encompassing differences in complexity, credibility, and potential loss.

2. The Conventional Generation: Phishing (Exploiting Human Vulnerabilities)

Phishing represents the initial phase of organized cybercrime. The core of Phishing is social engineering, where attackers attempt to manipulate victims into voluntarily disclosing sensitive information or performing harmful actions.

b. Complexity: Phishing possesses relatively low technical complexity. Attacks often take the form of mass email blasts sent without specific targeting. Although variants like Spear Phishing (more targeted) emerged, the foundation remains the design of fake interfaces that mimic trusted entities (e e.g., banks or online services).
c. Credibility: The credibility of Phishing relies on the quality of visual mimicry and psychological pressure (sense of urgency). However, Phishing can often be detected through anomalies such as grammatical errors, mismatched sender email addresses, or suspicious links. Its credibility easily collapses upon careful examination.
d. Potential Loss: Although it can cause significant financial losses in the aggregate (from numerous victims), losses at the single individual or corporate level are often confined to credential theft or direct financial loss from small transfers.

Transition and Escalation (The Intermediate Generation)

Prior to Deepfake, an escalation occurred through more structured attacks and sophisticated malware, such as Ransomware and Supply Chain Attacks. This stage was characterized by an increase in technical complexity facilitated by Cybercrime as a Service (CaaS) (Manky, 2013), where advanced tools became readily accessible on the digital black market. Nevertheless, the initial point of attack (vector) often still rooted in Phishing or software vulnerability exploitation, rather than media content manipulation.

The New Generation: Deepfake (Exploiting Digital Reality)

Deepfake represents a quantum leap, driven by the integration of Artificial Intelligence (AI), specifically Generative Adversarial Networks (GANs) and Deep Learning(Goodfellow et al., 2020). These attacks are no longer about tricking victims into clicking a link, but about making victims believe information that is entirely false.

a. Complexity: The technical complexity of Deepfake is very high. It requires significant computational power and sophisticated AI models to generate seamless (flawless) audio and visual content. However, once Deepfake tools become available as open source or through CaaS, the technical complexity for the average perpetrator can drastically decrease, exponentially increasing the threat.
b. Credibility: This is the most fundamental difference. Deepfake's credibility approaches perfection. Deepfake creates false "digital evidence"(Casey, 2011) that appears convincing to human eyes and ears. This enables highly personalized and persuasive fraud, such as the replication of a CEO's voice (voice cloning) for emergency fund transfer authorization, or fake videos damaging the reputation of an official. Deepfake fundamentally compromises trust in what we see and hear (perceptual trust).
c. Potential Loss: The potential loss from Deepfake is catastrophic (Sharma et al., 2024). Beyond large instant financial losses (due to executive targeting), Deepfake inflicts immeasurable losses (intangible loss) such as:
    a. Permanently damaging a brand or individual's credibility.
    b. Manipulating political processes or stock markets through large-scale disinformation.
    c. Compromising identity security systems based on facial or voice biometrics.

This evolution demonstrates that the focus of cyber defense must shift from securing the endpoint and educating users about links, towards the validation of the authenticity of the information itself. This is the core of this research problem: understanding the scale of this threat change that necessitates a commensurate security and legal response.

To what extent are current cybersecurity strategies and technologies, predominantly focused on mitigating conventional threats (Phishing and malware), effective in addressing Deepfake and other emerging AI-based cyber threats, and what proactive adaptations are necessary within the digital security and identity verification frameworks to maintain the integrity of digital evidence and public trust

The second research problem critically evaluates the readiness of current cybersecurity frameworks to confront new AI threats like Deepfake, and identifies the proactive adaptations absolutely necessary to maintain the integrity of digital evidence and public trust.

Limitations of Conventional Defense Strategies

Most current cybersecurity strategies and investments are designed to combat conventional threats, specifically Phishing, malware, and software vulnerability exploitation. These defense tactics include:

a. User Education: Training employees to recognize suspicious links and attachments (anti-phishing training).
b. Network and Endpoint Defense: The use of firewalls, spam filters, and antivirus/anti-malware software.
c. Conventional Multi-Factor Authentication (MFA): Reliance on passwords and one-time codes (OTP) sent via text or app.

The Critical Gap (Ineffectiveness against Deepfake):

These strategies are proven ineffective against Deepfake because:

a. Failure of Perceptual Trust: Deepfake does not require victims to click links or download malware. Deepfake succeeds simply by convincing the victim that the visual or audio information they receive is authentic. User education cannot fully overcome this because human eyes and ears (or even simple facial/voice authentication systems) are easily fooled by AI-generated content.
b. Invalidation of Digital Evidence: Deepfake allows criminals to create false evidence (e.g., a recording of a CEO ordering a fund transfer). In a legal context, this complicates the establishment of authenticity (provenance) and the integrity of digital evidence, challenging established principles of proof in court.
c. Threat to Biometric MFA: When Deepfakes are sophisticated enough to realistically mimic a person's face or voice, less advanced biometric authentication systems become vulnerable, creating a security hole at what should be a strong point of identity verification.

Necessary Proactive Adaptations

To close the gap created by Deepfake and other AI-based threats, a paradigm shift is required from reactive defense (responding to known attacks) to proactive, integrity-based defense. These adaptations must be implemented within the security and digital identity verification frameworks:

Adaptations in the Digital Security Framework (Technical)

1. Investment in AI-based tools specifically trained to detect non-visual artifacts or anomalies in multimedia content generated by GANs.
2. Implementing identity verification systems that do not rely on a single mode (face or voice) but require a combination of liveness detection biometrics, behavioral analysis, and hardware- or blockchain-based authentication.
3. Utilizing technologies such as Distributed Ledger Technology (DLT) or blockchain to provide cryptographic timestamps and watermarks on original media, allowing recipients to verify the source and integrity of critical digital content.

Adaptations in the Legal and Policy Framework

1. Revision of Digital Evidence Regulation: Criminal and civil procedural law must revise the definitions and standards for admitting digital evidence to accommodate the possibility of AI manipulation, demanding stricter verification of origin.
2. Regulation of AI Use: A legal framework is needed to clearly regulate accountability and transparency regarding the creation and distribution of Deepfake content, especially when used for fraudulent or disinformation purposes.
3. Protection of Digital Reputation and Identity: The law needs to provide stronger protection and fast recovery mechanisms for Deepfake victims whose reputations are damaged, given that the resulting losses are permanent and immediate.

Failure to adopt these proactive adaptations means that cyber defense will continue to operate using a 20th-century roadmap to fight 21st-century threats. The integrity of digital evidence and public trust, which are the foundations of economic transactions, governance, and democracy, will continue to erode.

Thus, this research problem highlights the legal and technical urgency to overhaul cyber defense strategies not only to capture conventional malware but also to authenticate digital reality itself, ensuring the validity of every interaction and piece of circulating information.

Conclusion and Recommendations

Conclusion

This study confirms that cybercrime has undergone a fundamental tactical evolution, shifting from conventional social engineering methods (Phishing) toward sophisticated Artificial Intelligence (AI)-based attacks capable of multimedia manipulation (Deepfake). This shift establishes fundamental differences across three main aspects:

Complexity and Credibility: While Phishing relies on volume and human vulnerabilities that easily collapse under scrutiny, Deepfake offers near-perfect credibility with high technical complexity. Deepfake no longer deceives through fake text or links, but through the creation of fabricated digital reality, fundamentally undermining human perceptual trust.

Potential Loss: The losses caused by Deepfake are catastrophic, far exceeding the aggregate financial losses of Phishing. Deepfake poses a systemic risk to the integrity of digital evidence, corporate reputation, and institutional stability (through large-scale disinformation).

Defense Gap: Current cybersecurity strategies and legal frameworks, which are still dominated by the mitigation of Phishing and malware (such as user education and spam filters), are proven ineffective in confronting AI-based threats. A critical gap exists between the speed of threat evolution and the lagging response of cyber defense and legal mechanisms.

Therefore, a paradigm shift is required from reactive defense to proactive, integrity-based defense to authenticate the authenticity of the information itself, rather than merely securing the endpoint.

Recommendations

Based on the conclusions regarding the existing technical and legal gaps, the following proactive recommendations are proposed:

Technical and Security Strategy Recommendations:

1. Organizations, particularly in the financial and governmental sectors, must invest in Deep Learning-based tools specifically designed to detect artifacts or anomalies in multimedia content generated by GANs and voice cloning.
2. Identity verification systems must transition from conventional MFA to systems that incorporate liveness detection and various biometric and behavioral modes to thwart Deepfake impersonation attempts.
3. The technology industry must promote the use of Distributed Ledger Technology (DLT) or blockchain to provide cryptographic timestamps and watermarks on original media, enabling third parties to verify the integrity and origin of critical digital evidence.

Legal and Policy Recommendations:

1. Governments and judicial bodies must revise the Code of Criminal and Civil Procedure to establish higher standards for the admissibility and verification of digital evidence, particularly audio and visual recordings, to account for potential AI manipulation.
2. Regulations are required to explicitly govern accountability and transparency concerning the creation and distribution of Deepfakes for fraudulent or disinformation purposes. This includes imposing severe sanctions for the use of AI that damages reputation and threatens national stability.
3. Closer international cooperation is needed in sharing Deepfake threat intelligence and harmonizing legal policies to confront Cybercrime as a Service (CaaS) syndicates operating across jurisdictions.

Acknowledgements

The author expresses profound gratitude and sincere appreciation for the support and contributions of various parties that have made the completion of this article possible.

Special acknowledgment is extended to the Faculty of Economics, Law, and Humanities, Ngudi Waluyo University, for the facility support provided throughout the research process.

We also convey our sincere thanks to Dr. Hani Irhamdessetya, S.H., M.H., for their invaluable guidance, critical input, and insights in analyzing the complexity of cybercrime tactical evolution from both technical and legal perspectives.

Your contribution in shaping the structured analysis of these next-generation cybersecurity challenges is highly significant.

Finally, it is hoped that the results of this research will provide a significant contribution to the development of adaptive cyber defense strategies and a responsive legal framework against the threat of Deepfake.

References

Akyazi, U., van Eeten, M., & Gañán, C. H. (2021). Measuring cybercrime as a service (caas) offerings in a cybercrime forum. Workshop on the Economics of Information Security, 1–15.

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 563060.

Casey, E. (2011). Digital evidence and computer crime: Forensic science, computers, and the internet. Academic press.

Collier, B., & Clayton, R. (2022). A "sophisticated attack"? innovation technical sophistication and creativity in the cybercrime ecosystem. 21st Workshop on the Economics of Information.

Cuganesan, S., & Lacey, D. (2011). Developments in public sector performance measurement: a project on producing return on investment metrics for law enforcement. Financial Accountability & Management, 27(4), 458–479.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144.

Manky, D. (2013). Cybercrime as a service: a very modern business. Computer Fraud & Security, 2013(6), 9–13.

Schmitt, M., & Flechais, I. (2024). Digital deception: Generative artificial intelligence in social engineering and phishing. Artificial Intelligence Review, 57(12), 324.

Sharma, P., Kumar, M., & Sharma, H. K. (2024). GAN-CNN ensemble: a robust deepfake detection model of social media images using minimized catastrophic forgetting and generative replay technique. Procedia Computer Science, 235, 948–960.

Wall, D. S. (2024). Cybercrime: The transformation of crime in the information age. John Wiley & Sons.

Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology Innovation Management Review, 9(11).