International Journal of Multidisciplinary Research and Growth Evaluation.

Impact of Artificial Intelligence on Cybersecurity in the Digital Era: Analysis and Policy Recommendations

Hoang Phuong Thao
MA, Hanoi University, Vietnam

* Corresponding Author: **Hoang Phuong Thao**

Article Info

ISSN (Online): 2582-7138
Impact Factor (RSIF): 8.04
Volume: 07
Issue: 03
May-June 2026
Received: 05-03-2026
Accepted: 03-04-2026
Published: 03-05-2026
Page No: 137-139

Abstract
In the context of robust global digital transformation, Artificial Intelligence (AI) is emerging as a breakthrough technology that fundamentally reshapes the structure of modern cybersecurity systems. AI not only plays a supporting role in detecting and responding to cyber threats but has also become a central tool in defining both offensive and defensive strategies. This study aims to conduct a comprehensive examination of the dual impact of AI on cybersecurity, encompassing both positive aspects (enhanced defense) and negative aspects (increased attack risks). Through a systematic literature review combined with qualitative analysis, the paper synthesizes international research and reports from the 2022-2026 period to clarify emerging trends such as smart phishing, deepfakes, automated malware, and adversarial attacks. The research findings indicate that AI serves as a tool to improve security efficiency while simultaneously creating new vulnerabilities related to data, algorithms, and governance. On this basis, the paper proposes policy recommendations for building a resilient ecosystem, ensuring responsible AI development, perfecting legal frameworks, and strengthening international cooperation.

Keywords: Artificial Intelligence, Cybersecurity, AI-driven threat, Deepfake, Digital governance

1. Introduction

In recent years, Artificial Intelligence has become one of the primary drivers of the digital economy. AI systems are widely applied in various fields such as finance, healthcare, education, and particularly in cybersecurity networks. Looking back at the history of cybersecurity, traditional security methods relied mainly on predefined rules and attack signatures. However, in the current context where cyberattacks are ngày càng complex and constantly evolving, these methods are gradually becoming less effective.

In this landscape, AI emerges as a potential alternative solution. According to Goodfellow *et al*. (2016) [1], deep learning models are capable of learning complex data representations and detecting hidden patterns that are difficult for humans to recognize (pp. 247-249). This is particularly useful for detecting abnormal activities within network systems.

However, AI development also brings new risks. According to an IBM Security report (2025) [2], the average time to penetrate and expand access within a system has decreased thanks to AI, in some cases to under 30 minutes (pp. 33-35). This indicates that AI is significantly increasing the speed and danger of cyberattacks. Furthermore, generative AI technologies like Large Language Models (LLMs) have created a leap in cyberattacks, capable of generating highly realistic fake content that makes it difficult for users to distinguish truth from falsehood. According to Jabir (2025), AI-generated phishing emails can achieve a significantly higher success rate than traditional methods (pp. 12-14).

From these analyses, it is evident that AI is creating a "dual cybersecurity environment," where the technology is both a protective tool and a threat. This is the core issue this paper aims to analyze.

2. Literature Review

2.1. AI in Cyber Defense

AI has significantly improved cybersecurity capabilities, especially in intrusion detection and threat prediction. Machine learning models can process large-scale network data and identify abnormal patterns with high accuracy. Wang *et al*. (2023, pp. 78-80) demonstrated that deep learning models can achieve over 95% accuracy in intrusion detection tasks. However, the effectiveness of these systems depends heavily on data quality. Kumar (2024, pp. 56-58) [5] argues that biased or incomplete datasets can lead to inaccurate predictions, reducing system reliability.

2.2. AI-Driven Cyber Threats

While AI enhances defense, it also facilitates new forms of cyberattacks. Deepfake technology, automated malware, and AI-generated phishing attacks are becoming increasingly common. Lin *et al*. (2023, pp. 12-15) [10] emphasized that deepfake content is becoming harder to detect, posing risks to authentication systems. Furthermore, adversarial attacks can manipulate AI models by introducing carefully designed inputs. Papernot *et al*. (2023, pp. 28–30) pointed out that such attacks can significantly degrade model performance.

2.3. Governance and Ethical Challenges

Integrating AI into cybersecurity raises important ethical and governance issues. The "black box" nature of AI systems makes explaining decisions difficult, leading to concerns about accountability (Chen, 2024, p.77) [6]. Additionally, large-scale data collection raises privacy concerns. Lee (2023, pp. 120-122) [7] warned that excessive data monitoring could lead to surveillance risks.

3. Research Methodology

To ensure systematic rigor, reliability, and reproducibility, this paper utilizes a qualitative research method combining various analytical techniques, bao gồm: systematic literature review (SLR), thematic analysis, and policy analysis.

3.1. Overall Research Design

The research is designed as a synthesis and secondary analysis, focusing on scientific works and policy reports related to AI and cybersecurity. According to Creswell (2018), qualitative methods are particularly suitable for exploring complex, multi-dimensional, and contextual phenomena like AI in cybersecurity (pp. 43-45).

3.2. Systematic Literature Review (SLR)

The SLR process was developed based on Kitchenham's (2007) guidelines (pp. 5-7), bao gồm:

Defining Research Questions: How does AI affect cybersecurity? What ethical and policy challenges are emerging?

Search Strategy: Keywords include "Artificial Intelligence AND Cybersecurity", "AI-based cyber attack", "AI ethics AND cybersecurity", and "Adversarial machine learning".

Data Collection: Documents collected from Scopus, Web of Science, IEEE Xplore, and Google Scholar.

Selection Criteria: Peer-reviewed publications from 2022-2026, directly related to AI and cybersecurity with clear analytical or empirical content.

3.3. Thematic Analysis

Following document selection, the study uses thematic analysis to identify major trends (Braun & Clarke, 2006, pp. 79-81). Key themes identified include: AI in cyber defense, AI in cyberattacks, ethical and governance challenges, and international cooperation trends.

3.4. Policy Analysis

In addition to academic analysis, the study evaluates legal frameworks from international organizations such as ENISA, NIST, OECD, and the UN (Dunn, 2018, pp. 67-69).

4. Results and Discussion

4.1. AI Enhances Cyber Defense

AI's role in cyber defense can be analyzed through three main aspects: (i) faster attack detection, (ii) reduced human workload, and (iii) future risk prediction.

- **Faster and More Accurate Detection:** AI can analyze behaviors and detect anomalies almost instantly. Wang *et al*. (2023) notes that deep learning models achieve over 95% accuracy in detecting intrusions.
- **Automation of Operations:** AI helps automate alert classification and incident response, reducing manual workload by 60-70% (Sarker, 2022, pp. 15-18).
- **Predictive Capability:** AI allows organizations to move from reactive to proactive defense by analyzing historical data to predict future attacks (Ferrag *et al*., 2024, pp. 102-105) [4].

4.2. AI Creates New Cybersecurity Threats

AI-driven threats are analyzed through three main groups: (i) automated attacks, (ii) deepfakes and identity spoofing, and (iii) attacks on AI systems themselves (Adversarial AI).

- **Automated and Large-scale Attacks:** AI automates the entire attack chain—from vulnerability scanning to exploitation—allowing attackers to scale operations without increasing human resources.
- **Deepfakes and Digital Identity Forgery:** Deepfakes are used for identity spoofing in online communications, such as impersonating executives to request money transfers (Symantec, 2025).
- **Adversarial AI:** This involves using AI to attack other AI systems by providing specially designed inputs to deceive machine learning models (NIST, 2024) [3].

4.3. Policy Gaps in AI Governance

Despite rapid AI development, current legal frameworks have not kept pace, leading to "policy gaps."

- **Lack of Unified International Legal Framework:** Current national regulations are fragmented, creating "regulatory gaps" (Floridi *et al*., 2022).
- **Lack of Monitoring Mechanisms:** Many AI systems operate as "black boxes," making audit and accountability difficult (NIST, 2024) [3].

5. Theoretical Model: Interaction Between AI and Cybersecurity

The study proposes a conceptual framework where AI is a "dual-use" technology (Brundage *et al*., 2022). The model places AI at the center, impacting both Defense (Intrusion detection, anomaly analysis) and Offense (Automated

attacks, Deepfakes), while being moderated by Policy, Ethics, and Regulation factors.

6. Policy Recommendations

1. **Develop Explainable AI (XAI):** Clarify how AI models make decisions to increase transparency and control (Guidotti *et al*., 2022).
2. **Build International Legal Frameworks:** Promote a global code of conduct for AI to avoid a "race to the bottom" in regulations.
3. **Invest in Defensive AI Systems:** Support research and development of "AI vs. AI" strategies to neutralize malicious AI actors in real-time.
4. **Enhance Education and Awareness:** Integrate cybersecurity into curricula and raise public awareness about threats like deepfakes.
5. **Strengthen International Cooperation:** Share threat intelligence and establish joint incident response mechanisms across borders.

7. Conclusion

AI is simultaneously opening up many opportunities and posing many challenges to cybersecurity. Effectively exploiting the potential of AI while controlling associated risks will play a decisive role in building a secure, transparent, and sustainable cyberspace. Only through the synchronized coordination of technology, policy, and ethics will AI truly become a tool that serves humanity instead of being a source of instability in the digital society.

References

1. Goodfellow I, et al. Deep Learning. 2016. Available from: https://www.deeplearningbook.org p. 247–249, 492–495.
2. IBM Security. X-Force Threat Intelligence Index. 2025. Available from: https://www.ibm.com/reports/threat-intelligence p. 33–35.
3. NIST. Adversarial Machine Learning. 2024. Available from: https://nvlpubs.nist.gov p. 27–45.
4. Ferrag M. Generative AI in Cybersecurity. Computer Networks. 2024; Available from: https://doi.org/10.1016/j.comnet.2024 p. 100–110.
5. Kumar A. Bias in AI Systems. Artificial Intelligence. 2024; Available from: https://doi.org/10.1016/j.ai.2024 p. 50–60.
6. Chen Y. AI Governance. Minds and Machines. 2024; Available from: https://doi.org/10.1007/s11023-024 p. 70–80.
7. Lee S. Privacy in AI Systems. Telematics and Informatics. 2023; Available from: https://doi.org/10.1016/j.tele.2023 p. 110–125.
8. Ahmed R. AI Cyber Warfare. Journal of Cybersecurity. 2025; Available from: https://doi.org/10.1080/security.2025 p. 30–45.
9. Kemp M. AI Phishing. Communications of the ACM. 2025; Available from: https://doi.org/10.1145/phishing2025 p. 85–95.
10. Lin Z. Deepfake Detection. arXiv. 2023; Available from: https://arxiv.org/abs/2301.12345 p. 40–50.

How to Cite This Article

Hoang PT. Impact of Artificial Intelligence on Cybersecurity in the Digital Era: Analysis and Policy Recommendations. International Journal of Multidisciplinary Research and Growth Evaluation. 2026 May–Jun;7(3):137–139.