Advances in Community Services Research

OPEN ACCESS

Volume 4, Issue 2 (2026). 56-67

e-ISSN: 2985-9778

DOI: https://doi.org/10.60079/acsr.v4i2.877

ACSR | RESEARCH ARTICLE

The Development of Artificial Intelligence Technology and Cyber Security Threats

Siti Sa'adah [1] Yulianah [2*] Neysha Putri Vaquitasari [3] Eka Septiana [4] Rafi Rasidin [5] Wanda Laksniyunita [6]

[1, 2*, 3, 4, 5, 6] Universitas Kebangsaan Republik Indonesia, Bandung, Indonesia

Email: ntisitii2@gmail.com, yulianah1288@gmail.com, pneysha8@gmail.com, ekasep.150@gmail.com, rafirsdn03@gmail.com

ARTICLE HISTORY

Submitted	:	June 02, 2026
Reviewed	:	June 07, 2026 June 12, 2026
Revised	:	June 14, 2026
Accepted	:	June 15, 2026
Published	:	June 16, 2026

Conflict of Interest Statement:
The author(s) declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ABSTRACT

Purpose: This study aims to examine the relationship between advances in Artificial Intelligence (AI) and the evolving cybersecurity landscape by identifying emerging AI-enabled threats, exploring AI-based defense mechanisms, and analyzing the challenges of implementing AI-driven cybersecurity systems.

Research Method: A structured literature review using a qualitative descriptive approach was conducted. Relevant studies published between 2021 and 2025 were systematically identified through major academic databases using predefined search keywords. The selected studies were screened using inclusion and exclusion criteria and analyzed through thematic content analysis.

Results and Discussion: The findings indicate that AI simultaneously functions as a catalyst for increasingly sophisticated cyber threats and as an enabler of advanced cybersecurity defenses. Three dominant themes emerged: the evolution of AI-driven threats, the application of AI in cybersecurity defense, and the technical and organizational challenges associated with AI implementation.

Implications: The study highlights the importance of integrating technological innovation, human oversight, and governance frameworks to strengthen cybersecurity resilience.

Originality: This review provides a holistic perspective by simultaneously examining AI-enabled threats, defensive applications, and implementation challenges within cybersecurity ecosystems.

Keywords: artificial intelligence; cybersecurity; AI-enabled threats; cyber defense; structured literature review.

1. Introduction

The Fourth Industrial Revolution has accelerated the integration of digital technologies into economic, social, and organizational activities. Among these technologies, Artificial Intelligence (AI) has emerged as a transformative force capable of enhancing operational efficiency, supporting decision-making, and fostering innovation across various sectors. According to Russell and Norvig (2020), AI encompasses technologies such as Machine Learning and Deep Learning that enable systems to learn from data, recognize patterns, and perform tasks with minimal human intervention. While these capabilities have generated substantial benefits in areas such as industrial automation and healthcare, they have also increased dependence on intelligent systems in critical infrastructures.

Advances in Community Services Research
Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

However, the rapid adoption of AI has introduced new cybersecurity challenges. Cybersecurity, which traditionally focused on mitigating manually executed attacks, now faces AI-enabled threats characterized by greater automation, adaptability, and scalability. Emerging threats such as deepfake-enabled social engineering, AI-generated phishing campaigns, and adaptive malware demonstrate how malicious actors can exploit AI technologies to enhance the sophistication and effectiveness of cyberattacks. At the same time, AI offers considerable potential for strengthening cybersecurity through anomaly detection, threat intelligence analysis, and automated incident response. This dual role positions AI as both a source of emerging cyber risks and a strategic instrument for cyber defense, highlighting the urgent need to understand its implications within the evolving cybersecurity landscape.

Recent studies have highlighted the growing importance of AI in both offensive and defensive cybersecurity contexts. On the defensive side, Krishnappa (2023) and colleagues demonstrate that AI-based security systems improve threat detection and response capabilities by analyzing large volumes of data, identifying anomalies, and facilitating real-time reactions beyond the capacity of traditional security approaches. Similarly, Hu *et al.,* (2021), one of the most influential contributions in this domain, comprehensively maps security threats across the AI lifecycle, from data collection to deployment, while also outlining corresponding countermeasures.

However, evidence also suggests that AI substantially enhances the capabilities of malicious actors. Chakraborty *et al.,* (2023) emphasize that cybercriminals increasingly exploit AI to develop sophisticated threats, including AI-powered malware and deepfake-enabled phishing attacks. Likewise, Kumar *et al.,* (2023) explicitly argue that AI and machine learning technologies simultaneously strengthen cybersecurity defenses while facilitating cyberattacks. Findings from more recent studies reinforce this paradoxical relationship. Arif *et al.,* (2024) document the emergence of AI-generated attacks involving automated phishing, ransomware, and deepfake technologies. In contrast, Bakhronkulova *et al.,* (2025) highlight the development of AI-assisted malware and automated phishing mechanisms that increasingly challenge traditional security controls.

Conversely, Khan *et al.,* (2025) emphasize the defensive potential of AI in detecting anomalies and automating response mechanisms. Thapaliya & Bokani (2024) report that machine learning and deep learning techniques facilitate real-time threat detection across extensive datasets, while Diop *et al.,* (2025) describe the effectiveness of AI-based approaches in mitigating zero-day attacks and distributed denial-of-service (DDoS) incidents. Despite these promising capabilities, existing studies reveal substantial barriers to the practical implementation of AI-driven cybersecurity solutions, including concerns about data quality, model transparency, interpretability, and algorithmic bias (Bakhronkulova *et al.,* 2025).

Although previous studies have extensively examined either the defensive applications of AI or the emergence of AI-enabled cyber threats, the existing literature remains fragmented in at least three important respects. First, relatively few studies synthesize offensive and defensive perspectives within a single analytical framework. Second, many reviews focus predominantly on describing AI applications without critically comparing the effectiveness, limitations, and implementation challenges associated with different approaches. Third, the rapidly evolving nature of AI-driven threats has created a gap between emerging attack techniques and the organizational readiness required to address them effectively. Consequently, there remains a need for a comprehensive review that systematically integrates evidence regarding AI's dual role within the cybersecurity landscape.

In response to these gaps, this study addresses the following research questions: (1) What are the major forms of AI-enabled cybersecurity threats identified in recent literature? (2) How has AI been

Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

utilized to enhance cybersecurity defense mechanisms? Moreover, (3) What challenges and strategic considerations influence the implementation of AI-driven cybersecurity solutions in organizational settings? By addressing these questions, this study aims to provide a more integrated understanding of the opportunities and risks associated with the increasing adoption of AI technologies in cybersecurity contexts.

The novelty of this study lies in its synthesis of contemporary evidence on both the offensive and defensive dimensions of AI in cybersecurity through a unified analytical perspective. Unlike previous studies that tend to focus exclusively on either threat development or defensive applications, this review critically examines the interactions among AI-driven attacks, AI-based mitigation strategies, and the practical constraints that affect their implementation. Therefore, this study contributes to a more comprehensive understanding of AI-enabled cybersecurity and provides insights that may help researchers, practitioners, and policymakers design more resilient digital security frameworks.

The remainder of this paper is organized as follows. The next section describes the methodological procedures used to conduct the literature review. The subsequent section presents and discusses the findings concerning AI-enabled threats, AI-driven defensive mechanisms, and implementation challenges identified across the reviewed studies. Finally, the paper concludes by summarizing the principal findings, outlining the study's contributions and limitations, and proposing directions for future research.

2. Literature Review and Hypothesis Development

2.1 Artificial Intelligence and the Evolution of Cybersecurity Threats

The rapid advancement of Artificial Intelligence (AI) has fundamentally transformed the cybersecurity threat landscape, shifting it from predominantly manual, opportunistic attacks to more automated, adaptive, and intelligent forms of cyber aggression. Unlike conventional cyberattacks, AI-enabled threats can learn from environmental conditions, optimize attack strategies, and operate at scales that exceed human capabilities. Guembe *et al.,* (2022) argue that the integration of AI into malicious activities has increased both the sophistication and complexity of cyberattacks, particularly through the emergence of AI-powered malware, intelligent phishing campaigns, and adversarial attacks targeting machine learning systems. Similarly, Khan *et al.,* (2024) emphasize that advancements in AI technologies have not only strengthened defensive capabilities but have also empowered threat actors to develop attacks that are more evasive, adaptive, and difficult to detect using traditional security mechanisms. This evolution reflects a paradigm shift in cybersecurity, in which threat actors increasingly leverage computational intelligence to automate reconnaissance, identify vulnerabilities, and continuously refine attack methods. Moreover, Achuthan *et al.,* (2024) highlight that the growing dependence on AI-driven systems has introduced new adversarial vulnerabilities, creating opportunities for attackers to manipulate intelligent models and exploit weaknesses embedded within digital infrastructures. Consequently, cybersecurity can no longer be understood solely through the lens of conventional threats, as the emergence of AI-enabled attacks requires a broader understanding of how intelligent technologies reshape the nature, scale, and operational dynamics of cyber risks.

Recent developments in generative AI have further accelerated the evolution of cybersecurity threats by enabling highly persuasive and personalized forms of social engineering. Schmitt & Flechais (2024) demonstrate that generative AI technologies, including deepfakes and synthetic voice

Advances in Community Services Research
Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

generation, have transformed cyber deception from technical exploitation to sophisticated psychological manipulation that can undermine trust and authenticity in digital interactions. In parallel, Jackson (2023) explains that machine learning has facilitated the evolution of phishing attacks from generic mass-distribution schemes into highly targeted spear-phishing campaigns tailored to individual victims. The proliferation of Large Language Models (LLMs) has also introduced novel attack vectors, including prompt injection, adversarial instructions, and AI-assisted phishing generation, thereby expanding the offensive capabilities available to cybercriminals (Ferrag *et al.,* 2025). Furthermore, Ofusori *et al.,* (2024) identify a growing trend toward intelligent malware and automated exploitation techniques that challenge the effectiveness of conventional security architectures and necessitate a fundamental reconsideration of existing defense strategies. Collectively, these studies indicate that the evolution of cybersecurity threats in the AI era extends beyond incremental technological advancement; rather, it represents a structural transformation in how cyberattacks are conceived, executed, and scaled. Understanding this transformation is therefore essential for anticipating future threat trajectories and informing the development of more resilient cybersecurity frameworks that can respond to increasingly intelligent adversaries.

2.2 Artificial Intelligence as a Cybersecurity Defense Mechanism

Artificial Intelligence (AI) has emerged as one of the most significant technological advancements in modern cybersecurity, enabling organizations to move beyond reactive security practices toward more proactive, adaptive, and intelligent defense mechanisms. The increasing complexity and volume of cyberattacks have exposed the limitations of traditional security approaches, particularly those relying on static rules and signature-based detection methods. In response, AI technologies such as machine learning, deep learning, and advanced analytics have been widely adopted to enhance threat detection, anomaly identification, and security monitoring capabilities. Salem *et al.,* (2024) emphasize that AI-driven detection techniques provide substantial improvements in identifying evolving cyber threats by analyzing large-scale datasets and recognizing hidden patterns that conventional security tools may overlook. Similarly, Mohamed (2025) argues that AI and machine learning have become essential components of contemporary cybersecurity architectures, supporting intrusion detection, malware classification, threat intelligence generation, and automated defense systems. The ability of AI to continuously learn from new data enables security systems to adapt to changing attack patterns and detect previously unknown threats. The growing sophistication of cyber threats has also prompted the adoption of more adaptive and intelligence-driven security frameworks, underscoring the need for advanced technologies to support proactive defense and real-time threat mitigation (Alam *et al.,* 2025). This capability is particularly important in addressing sophisticated attacks such as zero-day exploits, advanced persistent threats (APTs), and polymorphic malware, which frequently evade traditional defenses. Furthermore, Markevych & Dawson (2023) demonstrate that AI-enhanced intrusion detection systems significantly improve the identification of complex attack behaviors by leveraging behavioral analysis rather than relying solely on predefined attack signatures. These developments illustrate how AI has transformed cybersecurity from a largely reactive discipline into a dynamic and intelligence-driven defense ecosystem.

Beyond threat detection, AI also plays a critical role in strengthening cybersecurity through predictive analytics, automated response mechanisms, and enhanced situational awareness. Ajala (2024) highlights that AI-based anomaly detection and threat prediction systems enable organizations to

Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

identify potential security incidents before significant damage occurs, thereby supporting a more proactive cybersecurity posture. In addition, Afolalu & Tsoeu (2025) explain that AI enhances cyber defense through automated incident response capabilities, allowing organizations to rapidly contain threats, reduce response times, and minimize operational disruptions. Recent advancements have also introduced federated learning approaches that improve security while preserving data privacy. Hernandez-Ramos *et al.,* (2025) note that federated learning-based intrusion detection systems enable collaborative threat detection without centralizing sensitive information, thereby addressing critical concerns related to privacy and data governance. At the same time, the increasing reliance on AI-driven decision-making has raised concerns regarding transparency and trustworthiness. To address these challenges, Explainable Artificial Intelligence (XAI) has emerged as an important area of cybersecurity research. Khan *et al.,* (2025) argue that XAI-based intrusion detection systems improve the interpretability of security decisions, enabling analysts to understand the rationale behind threat classifications and enhancing confidence in automated security operations. Collectively, these studies indicate that AI functions not only as a tool for detecting cyber threats but also as a comprehensive defense mechanism that supports prediction, prevention, response, and decision-making. As cyber threats continue to evolve in sophistication and scale, AI-driven cybersecurity solutions are increasingly becoming a strategic necessity for building resilient and adaptive digital security infrastructures.

2.3 Challenges and Strategic Considerations in Implementing AI-Driven Cybersecurity

Despite the significant potential of Artificial Intelligence (AI) to enhance cybersecurity capabilities, its implementation presents a range of technical challenges that may limit effectiveness and organizational adoption. One of the most critical issues concerns the quality, availability, and reliability of data used to train AI models. Since AI-driven security systems rely heavily on large datasets, inaccurate, incomplete, or biased data can undermine detection accuracy and decision-making. Ling *et al.,* (2023) emphasize that challenges such as data quality, adversarial manipulation, scalability limitations, and integration with existing security infrastructures remain major obstacles to the successful deployment of AI-based cybersecurity solutions. In addition, the increasing sophistication of cyber threats has exposed vulnerabilities in AI systems themselves, particularly through adversarial attacks that manipulate model outputs and evade detection mechanisms. Another major concern is the lack of transparency in many AI models, often referred to as the "black-box" problem. Mendes & Rios (2023) argue that limited explainability reduces trust in automated security decisions, particularly in high-risk environments where analysts must justify threat assessments and response actions. Similarly, Nebati *et al.,* (2023) highlight that the development of Explainable Artificial Intelligence (XAI) has become a strategic priority in cybersecurity because interpretability is essential for validating security decisions and ensuring accountability. However, explainability alone does not guarantee improved security outcomes. Chung *et al.,* (2024) caution that excessive reliance on explanations generated by AI systems may create a false sense of security if those explanations are incomplete, inaccurate, or misunderstood. Furthermore, Saeed & Omlin (2023) identify additional challenges related to algorithmic bias, model validation, transparency, and the long-term sustainability of AI systems, indicating that technological advancement must be accompanied by rigorous governance and continuous evaluation.

Beyond technical considerations, implementing AI-driven cybersecurity requires comprehensive organizational, strategic, and regulatory preparedness. As AI increasingly influences security operations and risk management processes, organizations must address governance issues

Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

related to accountability, privacy protection, and compliance with evolving regulatory frameworks. Alam *et al.,* (2025) argue that AI-powered cybersecurity introduces new dimensions of cyber risk that cannot be effectively managed through traditional security governance models alone, necessitating integrated risk management frameworks capable of addressing both technological and organizational vulnerabilities. The rapid emergence of generative AI technologies further intensifies these concerns by introducing new ethical and operational challenges associated with automated content generation, decision-making autonomy, and information manipulation. According to Dwivedi *et al.,* (2023), organizations adopting AI must establish clear governance structures, accountability mechanisms, and ethical guidelines to ensure responsible deployment and reduce unintended consequences. In addition, successful implementation depends on organizational readiness, including the availability of skilled personnel, adequate technological infrastructure, and a culture that supports human–AI collaboration. These factors are particularly important because cybersecurity decisions often require contextual judgment that cannot be fully automated. Collectively, the literature suggests that the effectiveness of AI-driven cybersecurity is determined not solely by technological sophistication but also by organizations' ability to balance innovation with transparency, governance, risk management, and human oversight. Consequently, future cybersecurity strategies should emphasize not only the adoption of advanced AI technologies but also the development of robust institutional frameworks to ensure their responsible and sustainable use.

3. Research Method

This study uses a qualitative approach, employing descriptive and analytical methods, to explore the relationship between advances in Artificial Intelligence (AI) and the dynamics of cybersecurity threats. This study employs a structured literature review to systematically identify, evaluate, and synthesize scholarly evidence on AI-driven cybersecurity threats, AI-based defense mechanisms, and implementation challenges. The literature was collected primarily from peer-reviewed international publications indexed in Scopus, Web of Science, SpringerLink, ScienceDirect, Wiley Online Library, and Emerald Insight to ensure academic rigor and source credibility. The review focused on studies published between 2021 and 2025 to capture recent developments in AI-enabled cybersecurity technologies.

The search process employed combinations of the following keywords: "artificial intelligence AND cybersecurity", "AI-driven cyberattacks", "AI-powered malware", "adversarial machine learning", "deepfake attacks", "AI-enabled phishing", "intrusion detection systems", "AI-based cyber defense", and "explainable artificial intelligence in cybersecurity". The initial search identified 146 records across all databases. After duplicate removal (n = 28), 118 articles remained for title and abstract screening. Following application of the inclusion and exclusion criteria, 47 studies were assessed through full-text review, resulting in 21 eligible studies for the final analysis.

The inclusion criteria consisted of: (1) peer-reviewed journal articles; (2) publications written in English; (3) studies published between 2021 and 2025; and (4) studies directly addressing AI-enabled cyber threats, AI-based cybersecurity defense mechanisms, or implementation challenges related to AI-driven cybersecurity. Studies were excluded if they were editorials, conference abstracts, duplicate records, non-English publications, or lacked direct relevance to the research objectives. To ensure the

Advances in Community Services Research
Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

quality and reliability of the evidence, each study was evaluated based on source credibility, methodological clarity, relevance to the research questions, and contribution to the field.

Data analysis was conducted using content analysis techniques. The selected studies were coded and categorized into three thematic domains: (1) Artificial Intelligence and the Evolution of Cybersecurity Threats, (2) Artificial Intelligence as a Cybersecurity Defense Mechanism, and (3) Challenges and Strategic Considerations in Implementing AI-Driven Cybersecurity. Cross-study comparison and thematic synthesis were subsequently conducted to identify recurring themes, research trends, similarities, differences, and knowledge gaps within the existing literature.

4. Results and Discussion

4.1 Analysis Results

The development of Artificial Intelligence (AI) technology has triggered a paradigm shift in the cybersecurity landscape, where AI now functions as a dual-purpose technology that expands the digital attack surface. In this role, AI is not only used for operational efficiency but also becomes a highly dangerous offensive tool due to its ability to automate information gathering and target identification on a massive scale. Attackers can now use intelligent algorithms to map system vulnerabilities with high precision, thereby increasing the likelihood of successful exploitation across multiple digital infrastructures simultaneously.

The transformation from rigid manual attacks to adaptive, automated attacks marks a new era of cyber threats that are difficult to predict by conventional defense methods. Unlike the static code of the past, AI-based malware can learn to evade detection by traditional antivirus systems by changing its behavior and structure in real time. This adaptability allows threats to remain relevant even after defense systems are updated, creating a digital arms race where data processing speed is a key factor. The reviewed studies consistently indicate that AI increases the scalability and adaptability of cyberattacks, enabling threat actors to conduct sophisticated operations more efficiently than conventional attack methods.

The findings further reveal that AI facilitates the automation of reconnaissance and vulnerability identification, enhancing attackers' ability to exploit weaknesses across multiple digital infrastructures simultaneously. This scalability means a single threat actor can launch highly personalized and coordinated attacks across multiple sectors globally within a relatively short period. One key finding from this discussion is the emergence of increasingly sophisticated forms of social engineering through deepfake technology. By utilizing Generative Adversarial Networks (GANs), cybercriminals can create highly realistic audio and video manipulations that are difficult to distinguish from the original content. This technology allows for accurate imitation of a person's face and voice, making them potentially useful for breaching biometric-based security systems or committing identity fraud. As a result, verification methods previously considered secure, such as facial or voice recognition, are now more vulnerable to abuse. Furthermore, advances in artificial intelligence technology have also increased the effectiveness of phishing attacks. With the help of Natural Language Processing (NLP), perpetrators can craft fraudulent emails with neat, professional language and minimal errors. Furthermore, message content can also be personalized based on the victim's data, making it appear more convincing and relevant. This makes it more difficult for email recipients to distinguish between genuine and fraudulent messages, increasing the likelihood of successful attacks.

Advances in Community Services Research
OPEN ACCESS
Volume 4, Issue 2 (2026). 56-67
e-ISSN: 2985-9778
DOI: https://doi.org/10.60079/acsr.v4i2.877

The reviewed studies also suggest that organizations increasingly face challenges in maintaining information security due to the growing sophistication of AI-enabled attacks. The convergence of deepfakes, AI-assisted phishing, and intelligent malware has blurred traditional indicators for identifying malicious activity, thereby increasing organizational vulnerability. The emergence of AI-powered malware marks the declining effectiveness of traditional security systems that rely heavily on signature databases. This type of malware exhibits adaptive behavior that enables it to evade conventional detection mechanisms and maintain operational effectiveness despite updates to defensive systems. This AI-driven adaptability allows threat variants to evolve rapidly, challenging the sustainability of manual blacklist approaches.

The application of Adversarial Machine Learning techniques further exacerbates the sophistication of these attacks. In this scenario, attackers no longer hack code, but "hack" the logic of the defense AI model. By providing subtly manipulated data input, attackers can create "blind spots" in security systems. As a result, threat detection engines may classify malicious activities as normal data traffic or legitimate user behavior. In addition to evading detection, AI-based malware can execute highly personalized and scalable exploits. Once infiltrated, it can observe communication patterns and employee behavior within the network to determine the best time to exfiltrate data or launch ransomware encryption. This intelligent reconnaissance capability enables attackers to prolong their presence within targeted environments while minimizing the likelihood of triggering security alerts.

The implementation of AI in Security Operations Centers (SOCs) has transformed how organizations respond to threats, outpacing human capabilities. With the ability to process large volumes of network activity logs, machine learning can filter data noise to identify subtle attack patterns. This enables real-time anomaly detection, allowing security teams to intervene before systemic damage occurs and significantly reducing response times. However, heavy reliance on defensive AI introduces significant new risks, particularly the risk of systemic failure. If AI models that form the backbone of defense are successfully manipulated through adversarial attacks, the entire security infrastructure could lose its ability to distinguish between malicious and legitimate activities. Furthermore, algorithmic bias may contribute to inaccurate threat assessments, potentially resulting in either false positives or overlooked attacks.

4.2 Discussion

The findings suggest that Artificial Intelligence has fundamentally transformed the cybersecurity landscape by reshaping both the nature of cyber threats and the mechanisms used to counter them. Unlike conventional cyberattacks that rely heavily on predefined scripts and manual execution, AI-enabled threats exhibit automation, scalability, and adaptability, enabling attackers to optimize attack strategies and continuously adjust their methods in response to defensive measures. This transformation supports the argument presented by Guembe *et al.,* (2022) that AI has expanded the operational capabilities of threat actors through intelligent malware, automated phishing, and adversarial attacks. At the same time, the emergence of deepfake technologies and AI-assisted social engineering indicates that cybersecurity threats are no longer limited to technical vulnerabilities but increasingly exploit cognitive and behavioral dimensions of human decision-making (Schmitt *et al.,* 2024). This finding extends previous discussions by demonstrating that the evolution of cyber threats in the AI era involves both technological sophistication and psychological manipulation.

Advances in Community Services Research
Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

The reviewed studies also indicate that AI-driven defense mechanisms have become essential components of modern cybersecurity strategies. The integration of machine learning into Security Operations Centers (SOCs) enables organizations to process large volumes of security-related data more efficiently, improve anomaly detection, and accelerate incident response processes. These findings are consistent with Salem *et al.,* (2024) and Mohamed *et al.,* (2025), who argue that AI-based systems significantly enhance threat detection capabilities beyond those offered by traditional signature-based approaches. However, the effectiveness of AI-driven cybersecurity should not be interpreted as evidence that automation alone is sufficient. The literature consistently highlights that human expertise remains indispensable in interpreting contextual information, validating high-risk decisions, and responding to unprecedented attack scenarios that automated systems may not adequately address. Therefore, the relationship between humans and AI in cybersecurity should be viewed as complementary rather than substitutive.

Another important finding concerns the growing vulnerability of AI systems themselves. As organizations increasingly rely on intelligent defense mechanisms, attackers are developing strategies to exploit weaknesses in those systems, particularly through adversarial machine learning techniques. The findings support the arguments of Mendes and Rios (2023) and Srivastava *et al.,* (2022), who emphasize that transparency and explainability remain critical challenges in AI-based cybersecurity environments. The inability of analysts to understand how AI models generate security decisions may reduce trust in automated systems and increase organizational risk. Furthermore, excessive dependence on opaque algorithms may create conditions in which erroneous classifications remain undetected until significant damage has occurred. These findings suggest that cybersecurity resilience depends not only on algorithmic performance but also on the interpretability and trustworthiness of AI models.

The discussion further reveals that the successful implementation of AI-driven cybersecurity requires broader organizational and governance considerations. Technical sophistication alone does not guarantee effective cybersecurity outcomes if organizations lack adequate governance structures, skilled personnel, and mechanisms for continuous model evaluation. This observation aligns with Radanliev *et al.,* (2022), who argue that AI introduces new dimensions of cyber risk that necessitate revised governance frameworks and integrated risk management strategies. Similarly, Dwivedi *et al.,* (2023) emphasize the importance of accountability, ethical oversight, and regulatory preparedness in deploying intelligent technologies. Consequently, the future of cybersecurity should not be conceptualized solely as a technological competition between offensive and defensive AI capabilities. Rather, it should be understood as an ecosystem in which technological innovation, human expertise, ethical responsibility, and institutional preparedness collectively determine organizational resilience against increasingly sophisticated cyber threats.

5. Concluding Remarks and Recommendation

This study examined the evolving relationship between Artificial Intelligence (AI) and cybersecurity through a structured literature review focusing on three research areas: the evolution of AI-enabled cyber threats, the role of AI as a cybersecurity defense mechanism, and the challenges associated with implementing AI-driven cybersecurity systems. The findings indicate that AI has fundamentally transformed the cybersecurity landscape by simultaneously strengthening defensive capabilities and expanding the sophistication of cyber threats. The review demonstrates that AI-enabled attacks have

Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

become increasingly adaptive, automated, and scalable, as reflected in the emergence of intelligent malware, deepfake-assisted social engineering, adversarial machine learning, and highly personalized phishing campaigns. At the same time, AI-based cybersecurity solutions have enhanced threat detection, anomaly identification, and incident response capabilities, underscoring AI's dual role in shaping contemporary cybersecurity environments.

The study contributes theoretically by providing an integrated understanding of AI as both an offensive and defensive force within cybersecurity ecosystems. Unlike previous studies that predominantly focused on either AI-enabled threats or AI-based defense mechanisms, this review synthesizes both dimensions while incorporating implementation challenges related to explainability, governance, and organizational readiness. From a practical perspective, the findings emphasize the importance of balancing technological innovation with human oversight, organizational preparedness, and continuous model evaluation. From a policy standpoint, the study underscores the need for regulatory frameworks and ethical guidelines to support the responsible deployment of AI technologies in cybersecurity contexts. Consequently, the originality of this study lies in its holistic examination of AI-driven cybersecurity through a threat–defense–implementation perspective.

Several limitations should be acknowledged. First, this study relied exclusively on secondary data obtained from published literature, thereby excluding primary evidence from industry practitioners and cybersecurity professionals. Second, the review focused on studies published within a limited time frame, which may have omitted emerging developments after the literature search was completed. Third, despite employing a structured review approach, variations in methodological quality among the selected studies may influence the comprehensiveness of the synthesized findings. Future research should extend the current findings by incorporating empirical investigations involving cybersecurity experts, organizational case studies, and cross-sector analyses of AI implementation. Additional studies are also needed to evaluate the effectiveness of Human-in-the-Loop approaches, explainable AI frameworks, and governance mechanisms in real-world cybersecurity settings. Furthermore, as generative AI technologies continue to evolve, future research should explore their long-term implications for cyber resilience, regulatory development, and ethical decision-making in increasingly intelligent digital ecosystems.

Statement of Use of Generative AI

During the preparation of this work, the author used generative artificial intelligence tools to support the scientific writing process. Grammarly was used to check grammar, refine writing style, and improve clarity in scientific writing. All interpretations, analyses, and conclusions presented in this study are the sole responsibility of the author.

References

Achuthan, K., Ramanathan, S., Srinivas, S., & Raman, R. (2024). Advancing cybersecurity and privacy with artificial intelligence: current trends and future research directions. *Frontiers in Big Data*, *7*, 1497535. https://doi.org/10.3389/fdata.2024.1497535

Afolalu, O., & Tsoeu, M. S. (2025). Artificial Intelligence as the Next Frontier in Cyber Defense: Opportunities and Risks. *Electronics*, *14*(24), 4853. https://doi.org/10.3390/electronics14244853

Ajala, O. A. (2024). *Leveraging AI/ML for anomaly detection, threat prediction, and automated response.* https://doi.org/10.20944/preprints202401.0159.v1

Alam, M. A., Sarna, S. A., Rakibuzzaman, M., & Reza, J. (2025). Strengthening Cybersecurity Protocols to Safeguard

Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
OPEN ACCESS
e-ISSN: 2985-9778

U.S. Financial Infrastructure Against Emerging Threats. *Advances in Economics & Financial Studies*, *3*(2 SE-Articles), 71–82. https://doi.org/10.60079/aefs.v3i2.506

Alam, R. G. G., Hidayah, A. K., Gunawan, G., Wijaya, A., & Abdullah, D. (2025). *Manajemen Risiko Keamanan Informasi*. PT. Sonpedia Publishing Indonesia.

Arif, A., Khan, M. I., & Khan, A. R. A. (2024). An overview of cyber threats generated by AI. *International Journal of Multidisciplinary Sciences and Arts*, *3*(4), 67–76. https://doi.org/10.47709/ijmdsa.v3i4.4753

Bakhronkulova, L., Ali, M., Khabirova, Z., Azimjon, A., Zebo, A., & Muslima, A. (2025). Artificial Intelligence in Cybersecurity: Threats, Defenses, and Future Directions. *ENVIRONMENT. TECHNOLOGY. RESOURCES. Proceedings of the International Scientific and Practical Conference*, *2*, 23–30. https://doi.org/10.17770/etr2025vol2.8592

Chakraborty, A., Biswas, A., & Khan, A. K. (2023). *Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation BT - Artificial Intelligence for Societal Issues* (A. Biswas, V. B. Semwal, & D. Singh (eds.); pp. 3–25). Springer International Publishing. https://doi.org/10.1007/978-3-031-12419-8_1

Chung, N. C., Chung, H., Lee, H., Brocki, L., Chung, H., & Dyer, G. (2024). False sense of security in explainable artificial intelligence (XAI). *ArXiv Preprint ArXiv:2405.03820*. https://doi.org/10.48550/arXiv.2405.03820

Diop, M. M., Ba, M., Sylla, K., & Ouya, S. (2025). Artificial Intelligence in Cybersecurity: Applications, Challenges, and Future Developments. *2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, 1–6. https://doi.org/10.1109/IRASET64571.2025.11008137

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ferrag, M. A., Alwahedi, F., Battah, A., Cherif, B., Mechri, A., Tihanyi, N., Bisztray, T., & Debbah, M. (2025). Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities. *Internet of Things and Cyber-Physical Systems*, *5*, 1–46. https://doi.org/10.48550/arXiv.2405.12750

Guembe, B., Azeta, A., Misra, S., Osamor, V. C., Fernandez-Sanz, L., & Pospelova, V. (2022). The Emerging Threat of Ai-driven Cyber Attacks: A Review. *Applied Artificial Intelligence*, *36*(1), 2037254. https://doi.org/10.1080/08839514.2022.2037254

Hernandez-Ramos, J. L., Karopoulos, G., Chatzoglou, E., Kouliaridis, V., Marmol, E., Gonzalez-Vidal, A., & Kambourakis, G. (2025). Intrusion Detection based on Federated Learning: a systematic review. *ACM Computing Surveys*, *57*(12), 1–65. https://doi.org/10.48550/arXiv.2308.09522

Hu, Y., Kuang, W., Qin, Z., Li, K., Zhang, J., Gao, Y., Li, W., & Li, K. (2021). Artificial Intelligence Security: Threats and Countermeasures. *ACM Comput. Surv.*, *55*(1). https://doi.org/10.1145/3487890

Jackson, K. A. (2023). A systematic review of machine learning enabled phishing. *ArXiv Preprint ArXiv:2310.06998*. https://doi.org/10.48550/arXiv.2310.06998

Khan, M. I., Arif, A., & Khan, A. R. A. (2024). The most recent advances and uses of AI in cybersecurity. *BULLET: Jurnal Multidisiplin Ilmu*, *3*(4), 566–578.

Khan, M. I., Arif, A., & Khan, A. R. A. (2025). The Dual Role of Artificial Intelligence in Cybersecurity : Enhancing Defense and Navigating Challenges. *International Journal of Innovative Research in Computer Science and Technology (IJIRCST)*, *13*(1), 62–67. https://doi.org/10.55524/ijircst.2025.13.1.9

Khan, N., Ahmad, K., Al Tamimi, A., Alani, M. M., Bermak, A., & Khalil, I. (2025). Explainable AI-based intrusion detection systems for Industry 5.0 and adversarial XAI: A systematic review. *Information*, *16*(12), 1036. https://doi.org/10.3390/info16121036

Krishnappa, T. (2023). A Review on Artificial Intelligence Techniques in Preventing Cyber. *International Journal of Engineering Applied Sciences and Technology*, *8*(01), 185–189.

Kumar, N., Sen, A. C., Hordiichuk, V., Teresa, M., & Jaramillo, E. (2023). AI in Cybersecurity : Threat Detection and

Volume 4, Issue 2 (2026). 56-67
DOI: https://doi.org/10.60079/acsr.v4i2.877
e-ISSN: 2985-9778

Response with Machine Learning. *Tuijin Jishu/Journal of Propulsion Technology*, *44*(3), 38–46.

Ling, X., Wu, L., Zhang, J., Qu, Z., Deng, W., Chen, X., Qian, Y., Wu, C., Ji, S., Luo, T., Wu, J., & Wu, Y. (2023). Adversarial attacks against Windows PE malware detection: A survey of the state-of-the-art. *Computers & Security*, *128*, 103134. https://doi.org/https://doi.org/10.1016/j.cose.2023.103134

Markevych, M., & Dawson, M. (2023). A review of enhancing intrusion detection systems for cybersecurity using artificial intelligence (AI). *International Conference Knowledge-Based Organization*, *29*(3), 30–37. https://doi.org/10.2478/kbo-2023-0072

Mendes, C., & Rios, T. N. (2023). Explainable artificial intelligence and cybersecurity: A systematic literature review. *ArXiv Preprint ArXiv:2303.01259*. https://doi.org/10.48550/arXiv.2303.01259

Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems*, *67*(8), 6969–7055. https://doi.org/10.1007/s10115-025-02429-y

Nebati, E. E., Ayvaz, B., & Kusakci, A. O. (2023). Digital transformation in the defense industry: A maturity model combining SF-AHP and SF-TODIM approaches. *Applied Soft Computing*, *132*, 109896. https://doi.org/https://doi.org/10.1016/j.asoc.2022.109896

Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). Artificial Intelligence in Cybersecurity: A Comprehensive Review and Future Direction. *Applied Artificial Intelligence*, *38*(1), 2439609. https://doi.org/10.1080/08839514.2024.2439609

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*, 110273. https://doi.org/https://doi.org/10.1016/j.knosys.2023.110273

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. In *Journal of Big Data* (Vol. 11, Issue 1). Springer International Publishing. https://doi.org/10.1186/s40537-024-00957-y

Schmitt, M., & Flechais, I. (2024). Digital deception: generative artificial intelligence in social engineering and phishing. *Artificial Intelligence Review*, *57*(12), 324. https://doi.org/10.1007/s10462-024-10973-2

Thapaliya, S., & Bokani, A. (2024). Leveraging artificial intelligence for enhanced cybersecurity: Insights and innovations. *Sadgamaya*, *1*(1), 46–52. https://nepjol.info/index.php/sadgamaya/article/view/66888.

Corresponding author
Yulianah can be contacted at: yulianah1288@gmail.com