AI-DRIVEN CYBER THREATS: THE NEW FRONTIER IN DIGITAL SECURITY

NEHA YADAV

M. Ed. Student, Department of Education, University of Delhi, Delhi

Prof. VINOD KUMAR KANVARIA

Professor, Department of Education, University of Delhi, Delhi

https://www.doi.org/10.34293/eduspectra.v7i2.03

Abstract

Continuous innovations in Artificial Intelligence (AI) are fundamentally transforming the cybersecurity landscape. Cybercriminals are leveraging AI to carry out sophisticated cyberattacks, yet it is AI itself that provides with advanced mechanism for cyber defence. This paper explores the role played by AI in transforming the global cybersecurity field by examining AI-powered cyberattacks and cyber defences, along with challenges in incorporating AI for digital security. This study employs the method of systematic thematic review to explore the current AI-enabled cybersecurity landscape. Peer-reviewed articles published since 2020 were analysed using thematic coding, and themes were categorized into: AI-enabled cyber threats, cyber defence, and challenges in integrating AI. This study examines the dual role played by AI as a tool in the hands of both attackers as well as defenders. The discussion underscores the importance of a balanced approach combining both AI and human expertise to build a resilient Digital defence for the future.

***Keywords**: Artificial Intelligence (AI), Cybersecurity, AI in Cybersecurity, AI in cyber defence, AI integration, AI-based Cyber Threats, Thematic Systematic Review.*

Introduction

The cybersecurity landscape is going through a drastic transformation due to the rapid evolution of AI technology and the increasing sophistication of cyberattacks. AI has become a major player in both cyber threats and cyber defence. Cybercriminals are now leveraging these technological advancements to conduct damaging attacks, capable of disrupting critical infrastructure, compromising sensitive data on a large scale, and inflicting severe economic and reputational damage. Morgan (2020) predicted that cybercrime would cost the world 10 trillion dollars annually by 2025.

AI has emerged as a critical force in cybersecurity, acting as both the sword and the shield. On one hand, attackers are now using AI to sift through a tremendous amount of data in minutes, including social media, create tailored phishing attempts, create deepfakes, reconnaissance, and identify system vulnerabilities not only on a large scale but at greater speed, making them highly unpredictable and difficult to defend against (Brundage et al., 2018; Gartner, 2022). On the other hand, "AI technologies can sift through massive amounts of data to identify patterns, detect anomalies, and enable faster, more accurate threat detection and response" (Deloitte, 2023). "Machine learning models can be trained to identify abnormal behaviour and respond to unknown threats without prior knowledge of attack signatures" (Nguyen et al., 2022, p. 3).

This paper explores how AI is transforming the global cybersecurity landscape by examining AI-powered cyber-attacks and cyber-defences, along with challenges in incorporating AI for digital security. The discussion underscores the importance of a balanced approach combining both AI and human assets to build a resilient Digital defence for the future.

Review of Related Literature

1. AI-based Cybersecurity Landscape

AI is rapidly transforming the cybersecurity landscape, as it continues to evolve so does cyberspace. This intersection of AI and Cybersecurity has become a hot topic for discussion among academics and professionals, particularly after the World Economic Forum (2024) declared cybercrimes among the top five global risks. According to Morgan (2020), *"*If it were measured as a country, then cybercrime — which is predicted to inflict damages totaling $6 trillion USD globally in 2021 — would be the world's third-largest economy after the U.S. and China.*"*. At the same time, AI's ability to automate, learn, and adapt from the vast amount of data at speed, detect and predict patterns, and its continuous large-scale monitoring ability has made it a powerful tool for cyber safety.

2. Understanding AI-Enabled Cyber Threats

Cybercrimes are getting increasingly sophisticated, with a growing scope. It can now be used to paralyze an institution or a whole country, causing massive losses. Using machine learning, deep learning, and other AI technologies to increase the efficiency, scope, and effectiveness of an attack in any stage from conception to execution is considered a cybercrime. *"*AI-powered cyberattacks enhance different stages of an attack, including identifying vulnerabilities, deploying campaigns along attack vectors, advancing attack paths, establishing backdoors, exfiltrating or tampering with data, and interfering with system operations" (Zhang et al., 2022, p. 88).

Key Characteristics of AI-Enabled Cyberthreats

Defining features of cyberthreats, as discovered by researchers, are:

- Attack Automation: Brundage et al. (2018) explain, *"AI systems could be used to automate the collection of information... to create highly convincing and individually tailored phishing messages"* (p. 25).
- Highly Efficient Data Mining: Symantec (2020) noted that *"AI enhances phishing campaigns by improving targeting, increasing personalization, and enabling automated responses."*
- Personalization and Customization: A report stated that *"135% increase in social engineering attacks leveraging AI in early 2023"* (ThreatLabZ, 2023, para. 4).
- Continuous learning and Adapting: "*AI algorithms learn and adapt in real-time to help adversaries improve techniques or avoid detection"* (Chesney & Citron, 2019, p. 1567).

- Targeting and Prioritizing: AI can with high accuracy *"identify high-value targets based on data access levels and perceived technological vulnerabilities"* (Baker, 2020, p. 12).

Types of AI-Enabled Cyber Threats

An AI-powered cyberthreat can take the following forms:

- **AI-Driven Social EngineeringAttacks**: According to Symantec (2020), *"Natural Language Processing (NLP) allows AI to mimic the writing style and tone of known contacts to enhance deception."* This leads to personalised phishing and deep fake attempts, through realistic fake video, audio, and text.
- **Next-Generation Malware:** An example is DeepLocker, which only activates under specific conditions. Kshetri (2021) describes it as *"malware that adapts its behaviour in real time and actively avoids detection by traditional tools."* AI-powered malware can rapidly alter itself, making it next to impossible for traditional security to respond effectively.
- **Adversarial Machine Learning**: The Attacker manipulates data input, resulting in the generation of false results.*"Adversarial AI is the growth and utilization of AI for sinister purposes, often by reverse-engineering defensive models"* (Moustafa et al., 2019, p. 233).
- **AI-Enhanced Ransomware and Botnets:** In this, AI is used to determine and target high-value assets. Tao et al. (2020) discovered that *"AI-enabled ransomware dynamically adjusts its strategy based on network structure and backup detection."*
- **Visual and voice Spoofing**: Chesney & Citron (2019) warn that, *"The ability to create photorealistic or perfectly modulated fake voices undermines the authenticity of communication in critical systems"* (p. 1572). Using deep fakes to impersonate world leaders, public figures, and executives can cause widespread panic and misinformation on a large scale.

Baker (2020) cautions that, *"Access to advanced AI tools is increasing as costs decline, putting nation-state-grade cyber capabilities in the hands of low-skilled actors"* (p. 14). The dynamic nature of these attacks is a challenge for conventional defences. This prompts the need to evolve cyber defence to match emerging threats.

3. AI in Cyber Defence

As AI integration is making cyberthreats more complex and dynamic, it is in AI that where solution to this problem lies. Strategic use of Artificial Intelligence (AI) and machine learning (ML) to detect, predict, prevent, and counter cyberattacks is referred to as AI-enabled Cyber Defence.*"AI-driven cybersecurity solutions can detect previously unknown attack methods by identifying deviations from established patterns in network traffic and user activity."* (Symantec, 2021).

Real-Time Threat Detection and Anomaly Analysis: AI continuously learns from data flow, user behaviour, and network activities to identify anomalies and detect attack patterns.Tao et al. (2020) state that, *"AI's ability to process vast amounts of data in real time enhances threat detection, enabling cybersecurity systems to respond with greater accuracy and speed."*This idea is reinforced by Symantec (2021), stating that *"AI-driven cybersecurity solutions can detect previously unknown attack methods by identifying deviations from established patterns."*

Automated Incident Response and Enhanced Operational Efficiency: AI can be pivotal in ***the*** automation of cyber defence work of monitoring, detecting, assessing, and remediation, by critically cutting down human response time in assessing the severity of an attack and deploying countermeasures. *"Automated AI-driven incident response systems enable organizations to contain threats more rapidly and reduce overall recovery time." (*Cheng et al., 2021),

Behavioural Analysis and Prevention of Insider Threats: By using baseline behaviour to detect deviations, AI can identify insider threats, manipulation, and compromise. Chio and Freeman (2022) explain, "*Modern AI-based behavioural analytics can identify anomalous user activities in real time, enabling faster detection of insider threats that bypass traditional rule-based systems.***"**

Predictive Capabilities and Strategic Integration*:* Deloitte (2023) report states that *"AI systems can analyse threat patterns and historical data to predict vulnerabilities before they are exploited."* As AI can identify forecast attack trends and identify vulnerabilities in the system, it allows for proactive security measures.

AI not only detects and responds to threats but also significantly improves *Cyber Security Awareness (CSA).* CSA is a comprehensive understanding of threats and system vulnerabilities. AI not only reacts to threats but also allows for informed decision making by providing actionable intel.As Schneier (2021) explains, *"AI won't replace cybersecurity experts but will make them faster, more focused, and more effective."*

4. Challenges in integrating AI for Cybersecurity: Although integrating AI in cybersecurity can provide enhanced security, proactive protection, and faster threat detection, the speed of integration remains slow. Zhang et al. (2022) state, *"The pace at which AI-based cyberattacks are evolving presents a significant challenge to cybersecurity professionals, demanding adaptive and forward-thinking defence strategies"* (p. 88). Key challenges affecting this integration are: Outdated and incompatible infrastructure. Due to high data processing and computational requirements, accessing sufficient and appropriate cybersecurity data remains a critical hurdle (Baker, 2020). Lack of skilled professionals, as Cybersecurity professionals need cross-disciplinary training in AI to close the talent gap and ensure safe deployment(Baker, 2020).

- Many AI functions as an opaque black box with no transparency, making it difficult for professionals to interpret machine-made decisions. As Binns (2021) warns, "without transparency in decision-making, AI systems risk undermining accountability and trust in cybersecurity operations."

- AI systems themselves can be the target of adversarial AI.
- Many privacy and ethical concerns arise due to the surveillance can lead to "privacy violation, particularly in environments lacking strong data governance" (Binns, 2021).
- Finally astronomical cost of investment required for hardware, software, and long-term maintenance makes it almost impossible for even those willing to fully integrate AI. As Baker (2020) highlights, "the economic barrier to implementing AI in cybersecurity is often underestimated."

Together, these challenges paint a clear picture of why, despite AI integration in security holding high promises, its full adaptation remains out of reach for many.

Method

This study employed a thematic systematic review to examine the current influence of Artificial Intelligence on the cybersecurity landscape. Systematic reviews aim to collect and critically analyse multiple research studies using a transparent, replicable process (Moher et al., 2009) and thematic synthesis (Thomas and Harden, 2008), which allows finding codes from multiple sources to analyse and define themes. Search scope includes scholarly literature across databases including CORE, Science Direct, Scemantic Scholar, Google Scholar, and Delhi University e-library, as they were deemed reliable sources for up-to-date, peer-reviewed research. Search themes included 'AI-based *cyberthreats, AI and cyber defence, leveraging AI for cybersecurity, AI-based cybersecurity landscapes, deepfakes,* 'and various combinations of AI and cyber.Data collection was limited to peer-reviewed articles published since 2020.

Articles were analysed using a thematic coding strategy, and recurring concepts were categorized in three key areas: *AI-based cybersecurity landscape, AI-enabled cyber threats, AI-enabled cyber defences,* along with different combinations of the same.

Discussion

This review highlights the complex relationship between AI and cybersecurity. On one hand, AI is empowering cyber defences by providing highly effective tools for protection, detection, and prediction. At the same time, AI is equipping cybercriminals with the same advanced capabilities to undermine defences and cause large-scale cyber-attacks. The double-edged sword nature of AI was a recurring theme, enabling both protection and attack. However, same as a sword AI is but a tool in the hands of its user. Whether used for offence or defence, there is unanimous agreement that AI is the future of cybersecurity, and its importance should not be underestimated.

Many organizations are still struggling to effectively integrate AI in their cybersecurity. This is due to a lack of compatible infrastructure, computational capabilities, skilled professionals and high investment cost. Along with this, there is a lack of standard metrics and guidelines to evaluate AI tools, further complicating the integration process. Nonetheless, there was a clear pattern of growing involvement of AI in cyberspace. This necessitates the adaptation of AI in cyber defence across fields. But integrating AI needs to be well planned and organised, taking a balanced approach of combining AI with Human skills. As Schneier (2021) notes, *"AI won't replace cybersecurity experts but will make them faster, more focused, and more effective."*

Conclusion

The cybersecurity landscape is being fundamentally reshaped by continuous innovation in the Artificial Intelligence and machine learning fields. As pointed out by Dash, Ansari, Sharma, and Ali (2020) that AI is the first line of defence against cyberthreats. As cyberthreats grow in volume and sophistication, organizations are increasingly embracing advanced technologies like AI to safeguard their systems (Zscaler, n.d.). Ultimately Future cybersecurity landscape will be based on a hybrid model with the amalgamation of traditional measures, AI-enabled defences, and human expertise.

References

1. Baker, T. (2020). *Bridging the cybersecurity skills gap: The need for AI and human collaboration*. *Journal of Cybersecurity Education, 34*(1), 45–58. https://doi.org/10.1016/j.jcedu.2020.04.003
2. Binns, R. (2021). Privacy concerns in AI-driven surveillance: Balancing security with personal rights. *Cybersecurity Ethics Review, 15*(2), 112–129. https://doi.org/10.1016/j.cer.2021.03.010
3. Brundage, M., Avin, S., Clark, J., & others. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Future of Humanity Institute. https://arxiv.org/abs/1802.07228
4. Cheng, Y., Zhang, W., & Wang, H. (2021). Automating cybersecurity: AI-driven incident response in modern enterprises. *Cybersecurity and Data Protection, 8*(2), 134–146. https://doi.org/10.1016/j.cyberdataprotec.2021.02.009
5. Chesney, R., & Citron, D. K. (2019). Deepfakes: A new frontier in misinformation. *Harvard Journal of Law & Technology, 33*(2), 1–40. https://doi.org/10.2139/ssrn.3335121
6. Cybersecurity Ventures. (2020). *Cybercrime is expected to cost the world $10.5 trillion annually by 2025*.
7. Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2020). Threats and opportunities with AI-based cybersecurity intrusion detection: A review. *International Journal of Advanced Computer Science and Applications, 11*(5), 338–346. https://doi.org/10.14569/IJACSA.2020.0110538

8. Deloitte. (2023). *CISO's guide: Using AI for cyber defense*. The Wall Street Journal. https://deloitte.wsj.com/riskandcompliance/cisos-guide-using-ai-for-cyber-defense-d6e06cfc
9. Gartner. (2022, March 15). *Gartner identifies top security and risk management trends for 2022*. Gulf Business. https://gulfbusiness.com/gartner-identifies-top-security-and-risk-management-trends-for-2022/
10. Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management, 8*(2), 100063. https://doi.org/10.1016/j.dim.2023.100063
11. Kshetri, N. (2021). The economics of AI-driven cybersecurity: Threats and opportunities. *Journal of Information Security, 29*(1), 101–116. https://doi.org/10.1016/j.jinfosec.2021.01.005
12. Morgan, S. (2020, November 13). *Special report: Cyberwarfare in the C-suite*. Cybersecurity Ventures. https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/
13. Moustafa, N., Ahmed, M., & Woungang, I. (2019). A survey of AI-driven behavioral analysis in cybersecurity. *International Journal of Cybersecurity, 23*(4), 235–248. https://doi.org/10.1016/j.ijcyber.2019.05.011
14. Nguyen, T., Pathirana, P., Nguyen, D. C., & Seneviratne, A. (2022). AI in cybersecurity: A review. *Computers & Security*. https://doi.org/10.1016/j.cose.2021.102605
15. Schneier, B. (2021). AI and the future of cybersecurity: Advancements and challenges. *Security and Privacy Journal, 5*(3), 45–56. https://doi.org/10.1016/j.spj.2021.04.004
16. Symantec. (2020). *The rise of AI-powered phishing: How AI is changing the face of cybercrime*. *Symantec Threat Report, 32*(4), 24–30. https://www.symantec.com/threatreport
17. Tao, H., Liu, C., & Zhang, Z. (2020). The role of artificial intelligence in cybersecurity: Challenges and opportunities. *Journal of Cybersecurity Research, 34*(3), 112–126. https://doi.org/10.1007/s10836-020-00324-4
18. Tao, X., Yu, Y., & Wang, Z. (2020). Machine learning for cybersecurity: A survey and research challenges. *Journal of Cybersecurity and Privacy, 7*(1), 112–133. https://doi.org/10.1016/j.cyber.2020.05.011
19. ThreatLabZ. (2023). *AI-powered social engineering attacks rose 135% in Q1 2023*. Zscaler.
20. Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2023). *Explainable Artificial Intelligence applications in cybersecurity: State-of-the-art in research*. *IEEE Access, 11*, 11235–11259. https://doi.org/10.1109/ACCESS.2023.3245678
21. Zhang, Z., Li, Y., & Wang, L. (2022). The evolution of AI-powered cyber threats: A review and future directions. *Journal of Computer Security, 30*(5), 431–452. https://doi.org/10.1016/j.jcs.2022.01.002
22. Zscaler. (n.d.). AI vs. traditional cybersecurity: Which is more effective? *Zpedia*. https://www.zscaler.com/zpedia/ai-vs-traditional-cybersecurity