Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260502035
Volume 5, Issue2 (May-August 2026), PP: 318-326.
www.indjcst.com
ISSN No: 2583-5300

How Artificial Intelligence is Revolutionising Cybersecurity: Challenges and the Future

Santosh Kumar Jadala
Department of Computer and Information Sciences, University of the Cumberlands, United States of America.

To Cite this Article: *Santosh Kumar Jadala, "How Artificial Intelligence is Revolutionising Cybersecurity: Challenges and the Future", Indian Journal of Computer Science and Technology, Volume 05, Issue 02 (May-August 2026), PP: 318-326.*

Abstract: *The use of artificial intelligence (AI) in cybersecurity is growing to enhance threat detection, automate responses, enhance vulnerability management and support decision-making in a digital ecosystem. However, AI has also brought new challenges as the same tools that can be used to strengthen cybersecurity can also be used by cybercriminals to automate attacks, use smart malware, social engineering, adversarial attacks, and privacy breaches. In this article, we explore the double-edged effects of AI in cybersecurity through a review of the literature on AI-powered cyber defense, AI-powered cyber threats, organizational cybersecurity, ethical considerations and future research. This article argues that AI should not be seen as a substitute for human cybersecurity skills but as a sophisticated decision support and automation system that needs to be governed, transparent, accountable and monitored. The review finds that the potential of AI in cybersecurity will lie in the capacity of organizations, researchers and policymakers to manage innovation and risk, particularly in relation to adversarial AI, explainable AI, data protection, human accountability, and cyber governance.*

Key Word: *Artificial intelligence, cybersecurity, machine learning, cyber threats, cyber defence, adversarial AI, data protection, cybersecurity governance.*

I. INTRODUCTION

Cybersecurity is increasingly important to governments, organizations, corporations and individuals as digital technologies pervade all aspects of the economy, society and industry. The adoption of cloud computing, the Internet of Things, Industry 4.0, digital banking and other financial services, teleworking and distributed information systems has led to an increase in the value of digital information and the potential targets of cyber attacks. While conventional cybersecurity measures continue to be valuable, they are often inadequate in keeping up with the pace, volume and complexity of today's cyber attacks. This has led to the rise of artificial intelligence (AI) as a key technological innovation for future cybersecurity.

Artificial intelligence (AI) is the capacity of systems to carry out tasks that are usually performed by humans, involving learning, reasoning, pattern recognition, prediction, classification and decision-making. Machine learning, deep learning, natural language processing, expert systems and automated analytics are some of the common applications of AI in cybersecurity. Such technologies can be used to detect anomalous behaviour, malware, phishing, network traffic, vulnerabilities, and to assist with incident response. While some of the previous debates on AI and cybersecurity acknowledged the role of AI in supporting future security (Morel, 2011), more recent research has demonstrated that AI is now at the forefront of cybersecurity research and practice (Abbas et al., 2019; Naik et al., 2022; Wiafe et al., 2020).

But AI can be used as an offensive tool. It is increasingly being used by the bad guys. Hackers could exploit AI to automate intelligence gathering, craft and send realistic phishing emails, detect vulnerabilities in the system, evade detection or develop smart malware. This is a double-edged sword whereby AI can help the good guys but also the bad guys (Khan et al., 2024; Malatji & Tolah, 2025; Taddeo et al., 2019). As such, the effects of AI on cybersecurity need to be considered from both cyber defence and cyber attack vantage points.

This is a recent focus of authoritative sources. The new NIST Cybersecurity Framework 2.0 highlights the importance of governance as a cybersecurity function, particularly as companies implement AI-based systems (National Institute of Standards and Technology [NIST], 2024a). NIST's Generative AI Profile also points out the risks of misuse of data, vulnerability of models, unintentional generation of harmful content, and the need for a framework for managing AI-related risks (Autio et al., 2024). Likewise, ENISA Threat Landscape 2025 reveals that modern cyber threats are still prevalent and multifaceted, and the Verizon Data Breach Investigations Report 2025 illustrates the need to consider human factors, exploitations, and third-party risks in data breaches (European Union Agency for Cybersecurity [ENISA], 2025; Verizon, 2025). These reports indicate the need to address AI cybersecurity from the aspects of not just the technical, but also governance, risk and resilience. This study explores the benefits, risks and challenges of AI for cybersecurity. It is a narrative and conceptual review of the literature on AI applications, cybersecurity defense, enterprise cybersecurity, malicious AI, ethics and data privacy. The aim is to understand how AI will impact

cybersecurity, understand the benefits and challenges of AI-driven cybersecurity, and provide suggestions for future research and policies.

Research Questions

RQ1: How is AI enhancing threat detection and response?
RQ2: How are cybercriminals exploiting AI?
RQ3: What are the challenges of using AI-based cybersecurity solutions?

II. METHODOLOGICAL APPROACH

Review Design

This article employs a narrative literature review with a search strategy. A narrative review is suitable when the purpose is to integrate knowledge, draw general themes, and develop insights rather than statistically combine the answers to a research question. The multidisciplinary nature of the topic of AI and cybersecurity (spanning computer science, corporate risk management, ethics, privacy, and policy) means that a narrative approach enables the article to bring together literature from academic journals alongside select sources from cybersecurity authorities and standards bodies. The review is not intended to be a systematic review. But a search strategy was applied to enhance the transparency, relevance, and repeatability of the search process. The review uses six key areas of analysis: the use of AI in cybersecurity, AI-based cyber defense, AI-based cyberattack, ethical considerations and governance, organizational considerations and challenges, and future perspectives.

Search Sources

The literature was sourced from academic and authority/practice-based sources relating to AI and cybersecurity. Academic sources were: Google Scholar, IEEE Xplore, ScienceDirect, Springer Link, ACM Digital Library, MDPI, SSRN and publisher sources for the selected articles. Practical and authoritative sources were identified from reputable cybersecurity and standards bodies, such as NIST, ENISA, OWASP, CISA, IBM, Verizon and the UK National Cyber Security Centre.

Search Terms

The search terms used were variations on combinations of:
Search area, Example search terms:

- AI and cybersecurity, in general, "artificial intelligence" AND "cybersecurity"; "AI" and "cyber security"; "machine learning" and "cyber defense"
- Defensive AI: "AI threat detection"; "machine learning intrusion detection"; "AI malware detection"; "AI incident response"; "AI anomaly detection"
- Offensive and adversarial AI "adversarial AI" AND cyber security; "offensive AI"; "AI cyber attacks"; "AI phishing"; "AI malware evasion"
- Governance and ethics "AI cybersecurity governance"; (ethical AI) AND cybersecurity; (AI risk management); (explainable AI) AND cybersecurity
- Organizational cybersecurity: "AI effect on organizational cybersecurity"; "AI cybersecurity framework"; "AI cybersecurity readiness"
- Evidence of current threat: "data breach report," "cyber threat landscape, "generative AI cybersecurity risk, "LLM application security"

Inclusion and Exclusion Criteria

Sources were included if they were relevant to any of the following: AI applications in cybersecurity; use of AI-based cybersecurity; benefits or limitations of AI-based cybersecurity; adversarial or offensive AI; ethical, organizational and governance issues; AI in critical infrastructure or Industry 4.0; or current cybersecurity threat evidence from a trusted source. Preference was given to sources from 2017 onwards, but a single source by Morel (2011) was included due to its focus on early thinking on AI and cybersecurity. Sources were removed if they were not related to cybersecurity, concerned AI, but with no cybersecurity implications, did not contain adequate publication information, were a duplicate of another source, or presented non-scholarly opinion and did not offer any analysis. Industry and authority reports were only included if they presented current evidence or guidance for standards and cybersecurity practice.

Data Extraction and Synthesis

The chosen sources were thematically analyzed. The extracted information from each source included the paper's focus, its discussion of the AI technique or area in cybersecurity, the opportunities identified, the risks identified, the governance and/or ethical concerns, and the future research identified. These were distilled into broad themes of: AI for defense, AI for attack, limitations of AI, governance and ethics, readiness and future trends.

III. ARTIFICIAL INTELLIGENCE IN THE CYBERSECURITY CONTEXT

How AI relates to cybersecurity has changed over time. The initial focus of research was on AI as a potential tool to enhance cyber threat detection and response (Morel, 2011; Wirkuttis & Klein, 2017). Subsequent research broadened this conversation with a focus on the use of AI in malware detection, intrusion detection, security analytics, authentication, anomaly

detection and decision support (Abbas et al., 2019; Morovat & Panda, 2020; Wiafe et al., 2020). This is part of the evolution in cybersecurity from reactive to predictive and adaptive security approaches. AI-powered cybersecurity solutions are important because they can process more information faster than humans. Digital systems today produce large volumes of log data, network traffic, user activity logs, endpoint monitoring data, user authentication logs and system alerts. It can be challenging for humans to process this data. AI can help to identify patterns, anomalies and prioritize events for an investigation (Naik et al., 2022; Tao et al., 2021).

There are many applications of AI for cybersecurity. Abbas et al. (2019) explored the use of AI in cybersecurity and demonstrated that it has gained more interest. Wiafe et al. (2020) conducted a literature review on AI for cybersecurity and observed that AI has been used in various cybersecurity areas, including intrusion detection, malware detection, spam detection and network security. Likewise, Morovat and Panda (2020) stated that AI is a key tool to enhance cybersecurity due to its learning and detection capabilities. AI is particularly valuable in a dynamic environment. Signature-based security systems are based on previous knowledge of malicious behaviour. While these systems are not useless, they may not work if new types of attacks are used or if the malware is modified. AI-based systems can enhance detection by looking at anomalous behavior rather than just signature-based detection (Dalal, 2018; Naik et al., 2022). This is one way AI can be used to detect zero-day attacks, insider threats and advanced persistent threats.

IV. RESEARCH GAP

While the body of work on AI and cybersecurity is growing, a number of gaps can be observed. First, there is a strong emphasis on technological advantages of AI (including intrusion detection, malware identification and automated response) but less emphasis on organizational preparedness, governance, responsibility and human oversight. This results in a disconnect between technology and practice. Second, many studies talk about AI as a means of defence, but less about its use for offence and adversarial purposes. Recent research in adversarial and offensive AI has shown how AI may be leveraged by hackers to automate discovery, enhance evasion, and automate social engineering (Khan et al., 2024; Malatji & Tolah, 2025). In particular, more should be done in researching how to defend not only with AI but also against AI attacks.

Third, there's a lack of explainability. For example, an AI may do a good job of detecting threats, but cybersecurity professionals need to understand why the alert was triggered when making crucial decisions. It's not enough to get AI to be more accurate, but to make the AI outputs explainable, verifiable, and useful to analysts. Fourth, the field needs to better link academia and current cybersecurity practice. Authorities like NIST, ENISA, OWASP, CISA, IBM, Verizon, and the UK National Cyber Security Centre demonstrate that AI cybersecurity is now an issue that transcends technical research and is a matter of governance and resilience (Autio et al., 2024; ENISA, 2025; National Cyber Security Centre [NCSC], 2024; OWASP Foundation, 2025; Verizon, 2025). This article addresses these needs through a synthesis of academic and current authority sources and by considering AI cybersecurity from the technical, organizational, ethical and future perspectives.

V. AI'S OPPORTUNITIES IN CYBERSECURITY

Improved Threat Detection

Perhaps the biggest opportunity to be leveraged by AI is enhanced threat detection. AI can process network traffic, system logs, endpoint logs, and user logs and look for anomalies. Machine learning can be used to learn the difference between normal and abnormal activity, and can be used in intrusion detection systems (IDS) and anomaly detection systems (ADS) (Morovat & Panda, 2020; Naik et al., 2022). Security monitoring can generate a large number of (often false) alarms. AI can be used to prioritize alerts, eliminate false positives and correlate alerts. This will help cybersecurity analysts prioritize the threat. Lysenko et al. (2024) highlighted the importance of AI in providing protection and detecting threats, demonstrating how AI can enhance cybersecurity practices. Equally, Dalal (2018) suggested AI can help with threat detection and response by allowing systems to identify complex threats.

AI can also be used for behavioral analytics. Rather than just checking if an action has a "threat signature", AI can check for unusual behaviour of a user, device or application. For instance, if a user account suddenly starts downloading large amounts of data, signs in to the network from an unexpected location, or exhibits other activity outside of the normal workday, AI may detect the abnormal activity. This is crucial in preventing insider threats and compromised accounts.

Faster Incident Response

Second, there is a quicker incident response. Cybersecurity incidents sometimes need to be responded to quickly to limit damage, downtime and data loss. Automation can help with incident response by categorizing incidents, suggesting actions, quarantining systems, or initiating incident response procedures. This can be particularly beneficial in the case of a rapidly spreading attack. AI-enabled security systems can enhance an organization's ability to respond to cyber incidents in a timely fashion (Jonas et al., 2023). By using AI to assist in security operations centers, companies can enhance their ability to prioritize and analyze security incidents, identify patterns and respond accordingly. Calderon (2019) also noted the advantages of incorporating AI into cybersecurity, particularly with regard to automation and decision-making. But response automation must be done with caution. If an AI system responds to a threat incorrectly, it could lock out users, disable services, falsely flag normal user activity, and so on. Thus, AI response systems should be used in conjunction with human oversight, particularly in critical systems.

Malware Detection and Classification

Malware detection and classification have also become a possibility with AI. Malware is ever-evolving, and hackers often make changes in the malicious code to bypass traditional signature-based detection. Machine learning algorithms can analyze file properties, behaviours, system calls and code features to detect malware even without knowledge of the specific signature (Naik et al., 2022; Tao et al., 2021). AI-powered malware detection can detect ransomware, spyware, trojans, worms and other types of malware. It can also help with classification by clustering malware based on their characteristics and similarities with previous threats. This enhances threat intelligence and enables a quicker response. Morovat and Panda (2020) observed that AI can help cybersecurity operations by identifying patterns and classifying, which are crucial in malware analysis.

Phishing and Social Engineering Detection

Phishing is a popular type of cyber attack as it takes advantage of human vulnerabilities. AI can be used to identify phishing emails, suspicious websites, attachments and communication patterns. AI can be used to scan email content for fraudulent messages, detect manipulation and impersonation. AI can also detect patterns of communication that are out of the ordinary and may be an indicator of business email compromise and social engineering. But it can also be used to produce better quality phishing emails. As such, phishing is an example of the double-edged sword of AI in cybersecurity. Defenders use AI to identify phishing attacks, but so can attackers use AI to enhance the quality and personalization of their phishing (Khan et al., 2024; Malatji & Tolah, 2025).

Vulnerability Management and Predictive Security

AI can help in vulnerability management by assisting in the detection, prioritization and mitigation of vulnerabilities. Hundreds of vulnerabilities may be identified in large organizations in applications, networks and devices. Some vulnerabilities are more critical than others. AI can be used to prioritize vulnerabilities by exploitability, criticality of the asset, threat intelligence and past attack data. AI can also be used for predictive security. Rather than simply responding to attacks, AI can help predict and prevent potential attacks. AI can recognize these patterns, which may indicate a heightened risk, such as multiple failed logins, unusual scanning or network traffic. This allows a proactive approach to security.

Support for Industry 4.0 and IoT Security

AI is becoming more and more important in Industry 4.0, smart manufacturing and IoT. These systems are comprised of connected devices, sensors, machines and control systems. De Azambuja et al. (2023) discussed AI-based cybersecurity in Industry 4.0 and demonstrated that AI can help in protecting highly interconnected industrial systems. With increasingly digital industrial systems, AI can be used to monitor networks and anomalies in real time. IoT presents cybersecurity challenges because many of these devices are limited in processing capacity, have poor security by default and have long lifespans. AI can help by detecting device anomalies and compromised devices. But AI also adds complexity, as adversaries could also attack both the AI system and the systems it protects.

AI Benefits and Risks for Cybersecurity

The use of AI in cybersecurity has a dual-use nature. AI can help improve cybersecurity, but it can also be used by adversaries or create vulnerabilities. Table 1 shows the opportunities and risks identified from the literature.

AI opportunity	Cybersecurity benefit	Related risk or limitation
Threat detection	Improves identification of suspicious patterns, anomalies, and possible intrusions	May produce false positives or false negatives if data quality is poor
Malware analysis	Supports classification of malware families and detection of unknown variants	Attackers may modify malware to evade AI-based detection.
Phishing detection	Uses natural language processing and behavioral analysis to identify suspicious communication	Generative AI can help attackers create more convincing phishing content.
Automated incident response	Speeds up containment and reduces analyst workload	Incorrect automation may disrupt legitimate activity or critical services
Vulnerability prioritization	Helps rank vulnerabilities based on exploitability and asset importance	Model errors may misprioritize risks or overlook contextual factors
Threat intelligence	Processes large volumes of structured and unstructured threat data	Poor or manipulated intelligence data may weaken model reliability
Industry 4.0 and IoT monitoring	Detects abnormal behavior in connected devices and industrial systems	AI tools themselves may become targets in cyber-physical environments

AI opportunity	Cybersecurity benefit	Related risk or limitation
User behavior analytics	Identifies insider threats and compromised accounts	Excessive monitoring may create privacy and surveillance concerns
Generative AI security support	Assists analysts in summarizing alerts, drafting reports, and reviewing code	LLM applications may create risks such as prompt injection, insecure outputs, and sensitive information disclosure (OWASP Foundation, 2025)
AI governance and risk management	Supports structured cybersecurity and AI-risk decision-making	Weak governance may lead to shadow AI, unclear accountability, and compliance gaps

Table 1 Comparison of AI Opportunities and Risks in Cybersecurity

VI. CHALLENGES AND RISKS OF AI IN CYBERSECURITY

AI for Cyber Attackers

While AI enhances cybersecurity, it can also be used maliciously. This is a major concern. Khan et al. (2024) referred to AI as the new frontier of cybersecurity because AI can be used by both good and bad guys. Malatji and Tolah (2025) presented a model of adversarial and offensive AI, which demonstrated that AI can be used to enhance attack strategies, automation, evasion and exploitation. AI can be used to automatically scan for targets, discover new vulnerabilities, generate phishing emails, produce deepfake videos, enhance password cracking, bypass intrusion detection systems, or even create self-evolving malware. AI can also reduce the technical skill required for cybercriminals by enabling less-capable hackers to create scripts or social engineering emails. This doesn't mean that AI will necessarily make powerful attackers, but it can make existing cybercriminals more powerful. Attackers' use of AI results in a cybersecurity arms race. When defenders use AI-based tools, attackers may find ways to fool, trick or circumvent these tools. This makes adversarial testing and security for models more important.

Adversarial Attacks Against AI Systems

AI systems can be targets. Adversarial attacks are when data or inputs are manipulated to fool AI systems into making wrong predictions. For instance, cybercriminals can change the features of malware to evade detection or make network traffic look legitimate. They may also inject data into the learning process to teach a model bad behaviors. This threat is particularly significant because data quality is often critical for cybersecurity AI systems. Poor-quality data may be insufficient, biased, outdated, or even poisoned. Ansari et al. (2022) emphasized that there are limitations to the application of AI for cybersecurity, including issues with data integrity, model validity, and implementation. Similarly, Taddeo et al. (2019) warned that using AI for cybersecurity is a double-edged sword, as there is a risk of reliance on AI.

False Positives, False Negatives, and Model Reliability

AI cybersecurity tools are not infallible. False positives are when legitimate events are flagged as threats. False negatives are when threats are missed. Both can impact cybersecurity. False positives can cause false alarms, and false negatives can be missed. The model's effectiveness can be impacted by the data, algorithms, environment, and refreshing the model. Cyber threats are constantly changing, and if the AI model is not updated, it may not be as accurate. This is known as model drift. So cybersecurity models need to be periodically checked, trained, and validated. Likewise, Naik et al. (2022) noted that AI approaches can improve cybersecurity, but they have accuracy, adaptability, and complexity problems. Similarly, Namiot et al. (2022) observed that AI is related to cybersecurity, but AI designs and evaluations should be done correctly.

Lack of Explainability

AI systems, such as deep learning systems, can be black boxes. This can mean they may make a decision without explaining how they made their decision. Explainability may be vital to cybersecurity. Security analysts may need explanations of why a situation is suspicious. A lack of explanations can lead to analysts ignoring legitimate security events or falsely accepting suspicious events. Explainability is important for accountability. If an AI system blocks a user, blocks an operation, or misclassifies an event, the reason needs to be determined. A lack of explainability can cause operational and ethical issues for AI cybersecurity.

Privacy and Data Protection Concerns

AI systems require data, and cybersecurity systems may use or manage sensitive information such as network traffic logs, user behavior, access control rules, device and user information, and metadata. This can create issues of privacy and data protection. Zavushchak (2025) linked AI, cybersecurity, and data protection, noting that data collection, processing, storage, and data protection are crucial for using AI for security. Security can be improved with AI-based monitoring, but it can also compromise privacy. So it's important for organizations to consider security vs data protection. This means not collecting excessive data, protecting the data, controlling access, anonymizing where possible, and following policies and law.

Overreliance on AI

The second is overreliance. AI can help cybersecurity, but it can't decide, set policy, understand ethics, or strategize. According to Taddeo et al. (2019), reliance on AI in cybersecurity is a mixed blessing as AI can boost cybersecurity but also introduce new vulnerabilities. Organizations might be tempted to rely solely on AI to address cybersecurity issues, at the expense of training, governance, system hardening, and incident management. Cybersecurity is a socio-technical issue. It's about people, processes, technologies, policies, and culture. AI is just one element of this system. It needs to be integrated with other cybersecurity measures.

VII.ETHICAL AND GOVERNANCE IMPLICATIONS

Cybersecurity AI has implications for ethics in terms of responsibility, transparency, trust, privacy, fairness, and control. Taddeo (2019) discussed ethical issues with the use of AI for cybersecurity, such as the proper use of AI, accountability, and risk management. These are particularly relevant when AI is used to make decisions that impact users, employees, customers, or critical infrastructure. An ethical concern is accountability. If an AI-powered security system makes the wrong decision, is it the developer's, the company's, the cybersecurity analyst's, or the vendor's fault? This is more complex in the case of autonomous AI systems. Therefore, the organization needs to have governance arrangements with roles, responsibilities, and escalation to ensure the escalation of incorrect decisions.

Another issue is proportionality. Monitoring should be robust to ensure cybersecurity, but not overly intrusive to the point of privacy concerns and "Big Brother" surveillance. AI systems can monitor activities in bulk, which might help to identify threats, but could also cause some employees to be concerned. Trust is another important concern. Cybersecurity systems based on AI need to be trustworthy for security personnel, managers, and end users. But this does not imply blind trust. Taddeo et al. (2019) referred to trust in AI cybersecurity as a double-edged sword, as putting too much trust can be dangerous. Appropriate trust means testing, validation, explainability, auditing, and human oversight.

AI's Organisational Cybersecurity Impacts

AI impacts not just technical cybersecurity but also the entire cybersecurity strategy of an organization. Jada and Mayayise (2024) conducted a systematic review of the literature on the impact of AI on organizational cybersecurity and demonstrated that AI plays an important role in cybersecurity. AI can increase speed, automate processes, increase detection, and assist in decision-making. But it also demands new expertise, policies, and risk management. Companies using AI for cybersecurity need to be aware of some issues. First, they must have talent with expertise in cybersecurity and AI. Second, they need to ensure that AI solutions are integrated with their current security systems. Third, they need to work with data quality and privacy issues. Fourth, they need to be savvy about the vendors they choose, and not all AI is created equal.

Soni (2020) discussed challenges and prospects of AI in cybersecurity, noting that AI needs to overcome implementation challenges for success. These challenges are complexity, cost, skills, and data. Likewise, Sontan and Samuel (2024) highlighted opportunities and challenges for organizations with AI and cybersecurity. Organisations require a holistic approach to cybersecurity that includes AI. This involves using AI-based threat detection combined with firewalls, encryption, access control, endpoint security, security awareness training, incident response, vulnerability management and governance. AI should be integrated with an existing security infrastructure.

Cybersecurity with AI in Critical Industries

AI-driven cybersecurity is essential for critical sectors like energy, finance, health care, telecommunications, transportation, manufacturing and government. Cybersecurity threats in these industries can lead to financial, operational, safety, and national security problems. Basu (2024) penned an article on the use of AI for cybersecurity in the petroleum industry, showing that AI can be used in industrial sectors in which cybersecurity issues can affect their operations, technology and critical infrastructure. Cybersecurity issues in the industrial sector go beyond data theft. They can also cause damage to processes, equipment, or people. AI can help to monitor processes, detect anomalies, and predict failures. However, these systems need to be well managed as AI decisions can be critical. Cybersecurity is also a key element in Industry 4.0. De Azambuja et al. (2023) showed the importance of AI-based cybersecurity for smart manufacturing, protecting smart devices and systems. As AI is increasingly being used, cybersecurity should be a focus in AI governance.

VIII. CONCEPTUAL FRAMEWORK

In this article, we offer a framework of the impact of AI on cybersecurity. This composition relates to four factors: AI capability, cybersecurity application, risk, and governance.

Dimension	Explanation	Examples
AI capability	The technical functions that AI brings into cybersecurity	Machine learning, deep learning, natural language processing, automation, predictive analytics
Cybersecurity application	The defensive areas where AI is applied	Threat detection, malware analysis, phishing detection, vulnerability

Dimension	Explanation	Examples
		management, incident response, threat intelligence
Risk exposure	The new or intensified risks created by AI adoption and misuse	Adversarial attacks, AI-enabled phishing, model poisoning, false positives, privacy risks, and overreliance
Governance response	The controls needed to ensure responsible and effective AI use	Human oversight, explainability, data protection, model auditing, cybersecurity frameworks, and ethical policies

Figure 1 Conceptual Framework for AI Impact on Cybersecurity

The framework implies that cybersecurity outcomes are not solely dependent on AI capability. Rather, it is determined by the use of AI capabilities, the risks introduced by using (and misusing) AI, and the governance of those risks. For instance, machine learning can enhance malware detection, but its effectiveness is dependent on data integrity, robustness against adversaries, analyst understanding, and organizational responses. Likewise, generative AI may help security analysts by summarizing alerts or helping with code reviews, but it may introduce prompt injection, sensitive data leakage, and insecure code generation if not properly managed (Autio et al., 2024; OWASP Foundation, 2025).

The framework thus draws attention to the socio-technical nature of AI cybersecurity. Software performance needs to be considered in relation to human skills and training, organizational readiness, compliance and governance, ethical considerations, and risk management. This clearly points to the need for AI to be embedded into cybersecurity as part of a managed and multi-layered approach.

IX.FUTURE DIRECTIONS

Explainable AI for Cybersecurity

The key priority for future research is explainable AI. To support cybersecurity analysts, AI tools should not only indicate potential threats but also why they are potential threats. Explainability can increase trust, aid in investigation, minimize false alarms, and increase accountability. The next generation of AI cybersecurity systems should generate explanations that support a user's understanding of the model.

Adversarial AI and Model Security

Adversarial AI will be a growing area of research. With increasing AI use in cybersecurity, adversaries will likely attack the AI models. This will require research into how to defend AI models against poisoning attacks, evasion attacks, data manipulation, and model theft. Malatji and Tolah (2025) have called for a better understanding of adversarial and offensive AI, which should be a focus for future research.

Human-AI Collaboration

Human-AI collaboration is likely to be critical to the future of cybersecurity. AI may be able to handle big data, but humans have contextual knowledge, ethical decision-making, strategic planning, and innovation. Research should thus focus on how AI can aid analysts while leaving key decision-making to humans.

AI Governance and Regulation

AI governance will be crucial as it is used for security. Governance needs to cover issues such as accountability, transparency, privacy, procurement, audit, monitoring, and incident management. Polito and Pupillo (2024) have treated AI and cybersecurity from a policy viewpoint, demonstrating that cybersecurity has to be taken into account when dealing with AI.

Data Protection and Privacy-Preserving AI

Cybersecurity should also concentrate on privacy-preserving AI. AI relies on data, so researchers and practitioners need techniques to perform security analyses without compromising privacy. Methods such as anonymization, federated learning, differential privacy, and secure processing may be used more frequently.

Cybersecurity Education and Workforce Development

Cybersecurity is evolving due to AI. The next generation of cybersecurity professionals will need to know about AI, data analytics, model biases and attacks, and ethical considerations. Hence, companies and universities should add AI cybersecurity topics to their cybersecurity education programs.

X.RECOMMENDATIONS

The literature review allowed the identification of some recommendations. First, AI should be used as an aid, rather than a replacement for human cybersecurity professionals. AI can enhance detection and response, but humans are still required. Second, AI-driven cybersecurity should be regularly tested, updated, and audited. The threat landscape is ever-changing, and AI

models can degrade over time. Third, companies should use explainable AI. Cybersecurity analysts should know how and why AI models determine if an activity is malicious. Fourth, security analysts should be ready for adversarial AI. This involves evaluating AI models for evasion, securing data used to train AI models, and monitoring AI model performance. Fifth, security teams should include privacy and protection in their AI cybersecurity. Companies should collect only as much data as is necessary and ensure that their monitoring activities are legal, ethical, and justified by the benefits. Sixth, policymakers and academics should provide more guidance on the governance of AI in cybersecurity. Accountability, transparency, ethical considerations, and risk management are key aspects to be addressed. Lastly, cybersecurity training should cover AI topics to help professionals correctly use, assess, and manage AI cybersecurity tools.

XI. CONCLUSION

AI is revolutionizing cybersecurity by opening up new possibilities for threat detection, incident response, malware analysis, phishing, vulnerability management, and industrial cybersecurity. AI can assist organizations in analyzing vast amounts of security information, discovering subtle trends, and responding rapidly to incidents. Such capabilities are critical in a digital world where threats are common, sophisticated, and ever-changing. But AI also presents new challenges. The same capabilities that enhance cybersecurity can also be exploited by cyber criminals to automate and enhance social engineering, evade defences, and attack AI systems themselves. AI-powered cybersecurity solutions also exhibit issues with false positives, false negatives, data quality, explainability, privacy, and overreliance. Governance and ethical issues are therefore crucial for AI in cybersecurity.

Responsible integration will be key to the future of cybersecurity. Companies should view AI as an enabler but not a silver bullet in their cybersecurity strategy, which should also incorporate human, process, technology, governance, and learning. Research must look to explainable AI, adversarial AI, privacy-preserving security analytics, AI governance, and human-AI collaboration. If well-designed and managed, AI can improve cybersecurity. And if not managed well, it can create new risks and exacerbate existing ones. The challenge for the future is to use AI to make digital systems safer rather than to increase the capabilities of cyber adversaries.

Funding Statement
This research received no external funding

REFERENCES

1. Abbas, N. N., Ahmed, T., Shah, S. H. U., Omar, M., & Park, H. W. (2019). Investigating the applications of artificial intelligence in cybersecurity. Scientometrics, 121(2), 1189-1211. https://doi.org/10.1007/s11192-019-03222-9
2. Ali, A., Khan, M. A., Farid, K., Akbar, S. S., Ilyas, A., Ghazal, T. M., & Al Hamadi, H. (2023, March). The effect of artificial intelligence on cybersecurity. In the 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-7). IEEE. 10.1109/ICBATS57792.2023.10111151
3. Ansari, M. F., Dash, B., Sharma, P., & Yathiraju, N. (2022). The impact and limitations of artificial intelligence in cybersecurity: a literature review. International Journal of Advanced Research in Computer and Communication Engineering. https://ssrn.com/abstract=4323317
4. Basu, A. (2024, November). The impact of artificial intelligence on cybersecurity. In Abu Dhabi International Petroleum Exhibition and Conference (p. D021S077R001). SPE. https://doi.org/10.2118/222493-MS
5. Calderon, R. (2019). The benefits of artificial intelligence in cybersecurity. https://d1wqtxts1xzle7.cloudfront.net/104950055/viewcontent-libre.pdf?1691851508=&response-content-disposition=inline%3B+filename%3DThe_Benefits_of_Artificial_Intelligence.pdf&Expires=1777362488&Signature=F2gbQKxqgCeqCFqNFvJpb4a9d8Ii1xpVQEz8jXJq7FDEwgNR0KBAhGLys5yEVBf~AQOtTiwP8P~gWvz~3kQa3qvJPui6t9fmFi6UKYBOWaL3dZEZhsUH7LMUWFg7xcIdn6sQucz~NTZG7CpJwFWlzAGG05c3PnnRIDVWL6OVajdpRkUMdcsFEL~Eu5NkyLNh6rDegjzswCMM5OjoHs6Q-03KPwESUibXaaLEgn~ME1dks2f~7ZEi~yIcDQJbACYPXHQW2Z9nR1Ae8YMV3lbtEdFKjOcFKD6ZpucXBTZp0Of~QKVE2AGivBgv957t7GDg6cs3FtGHgoP9O36xYZqUg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA Dalal, A. (2018). Cybersecurity and artificial intelligence: How AI is being used in cybersecurity to improve detection and response to cyber threats. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 9(3), 10-61841. https://dx.doi.org/10.2139/ssrn.5403828
6. De Azambuja, A. J. G., Plesker, C., Schützer, K., Anderl, R., Schleich, B., & Almeida, V. R. (2023). Artificial intelligence-based cybersecurity in the context of Industry 4.0—a survey. Electronics, 12(8), 1920. **https://doi.org/10.3390/electronics12081920**
7. European Union Agency for Cybersecurity. (2025). ENISA threat landscape 2025. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025
8. IBM Security. (2025). Cost of a data breach report 2025. IBM. https://www.ibm.com/reports/data-breach
9. Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organizational cybersecurity: An outcome of a systematic literature review. Data and Information Management, 8(2), 100063. https://doi.org/10.1016/j.dim.2023.100063
10. Jimmy, F. (2021). Emerging threats: The latest cybersecurity risks and the role of artificial intelligence in enhancing cybersecurity defenses. Valley International Journal Digital Library, 1(2), 564-574. 10.18535/ijsrm/v9i2.ec01
11. Jonas, D., Yusuf, N. A., & Zahra, A. R. A. (2023). Enhancing security frameworks with artificial intelligence in cybersecurity. International Transactions on Education Technology (ITEE), 2(1), 83-91. https://doi.org/10.33050/itee.v2i1.428
12. Khan, K., Khurshid, A., & Cifuentes-Faura, J. (2024). Is artificial intelligence a new battleground for cybersecurity?. Internet of Things, 28, 101428. https://doi.org/10.1016/j.iot.2024.101428
13. Lysenko, S., Bobro, N., Korsunova, K., Vasylchyshyn, O., & Tatarchenko, Y. (2024). The role of artificial intelligence in cybersecurity: Automation of protection and detection of threats. Economic Affairs, 69, 43-51. https://www.proquest.com/openview/ebe74758eaa95b3e9ea9e2882b9cfb1d/1?pq-origsite=gscholar&cbl=2032164
14. Malatji, M., & Tolah, A. (2025). Artificial intelligence (AI) cybersecurity dimensions: a comprehensive framework for understanding adversarial and offensive AI. AI and Ethics, 5(2), 883-910. https://doi.org/10.1007/s43681-024-00427-4
15. Morel, B. (2011, October). Artificial intelligence and the future of cybersecurity. In Proceedings of the 4th ACM workshop on Security and

artificial intelligence (pp. 93-98). https://doi.org/10.1145/2046684.2046699
16. Morovat, K., & Panda, B. (2020, December). A survey of artificial intelligence in cybersecurity. In the 2020 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 109-115). IEEE. 10.1109/CSCI51800.2020.00026
17. Naik, B., Mehta, A., Yagnik, H., & Shah, M. (2022). The impacts of artificial intelligence techniques in the augmentation of cybersecurity: a comprehensive review. Complex & Intelligent Systems, 8(2), 1763-1780. https://doi.org/10.1007/s40747-021-00494-8
18. Namiot, D., Ilyushin, E., & Chizhov, I. (2022). Artificial intelligence and cybersecurity. International Journal of Open Information Technologies, 10(9), 135-147. https://injoit.org/index.php/j1/article/view/1402
19. National Cyber Security Centre. (2024). The near-term impact of AI on the cyber threat. https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat
20. National Institute of Standards and Technology. (2024a). The NIST cybersecurity framework (CSF) 2.0 (NIST CSWP 29). https://doi.org/10.6028/NIST.CSWP.29
21. Okdem, S., & Okdem, S. (2024). Artificial intelligence in cybersecurity: A review and a case study. Applied Sciences, 14(22), 10487. https://doi.org/10.3390/app142210487
22. OWASP Foundation. (2025). OWASP top 10 for large language model applications https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-v2025.pdf
23. Polito, C., & Pupillo, L. (2024). Artificial intelligence and cybersecurity. Intereconomics, 59(1), 10-13. 10.2478/ie-2024-0004
24. Shamiulla, A. M. (2019). Role of artificial intelligence in cybersecurity. International Journal of Innovative Technology and Exploring Engineering, 9(1), 4628-4630. https://d1wqtxts1xzle7.cloudfront.net/103482360/A6115119119-libre.pdf?1687044419=&response-content-disposition=inline%3B+filename%3DRole_of_Artificial_Intelligence_in_Cyber.pdf&Expires=1777363231&Signature=gWEMci4dKW8eGBw~DtSycE5jmUa9vJzKh94J0So2YIm0yjPVg4X0offutKePI8hiV28XAMcYMaLR6GhjQ7rbrQNUWSe6sUJK5dZ63XwU2hgvwhh0g8NCyZQSDyPxzWKpttftfREGyl3stgDh6hy9TgySR2oaPhv3CWJeJJHjA4Uu4xPkp2EtRPjcQsR4ATJjUOziwSp3Yqw355mJBu7iWfMT0aY6BHO7BJtY~65k5bUFxiOi2a3k~0xaY5qr44CLs9z3OoJKBtROCDAmW2dbgjmXPl3GNAtX9yAXj9PTQqz8RhW~lTr5RL3LtdU02n-YReRbvZ3r12ZBBGb8AZlrg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
25. Soni, V. D. (2020). Challenges and Solutions for Artificial Intelligence in Cybersecurity of the USA. Available at SSRN 3624487. https://dx.doi.org/10.2139/ssrn.3624487
26. Sontan, A. D., & Samuel, S. V. (2024). The intersection of Artificial Intelligence and cybersecurity: Challenges and opportunities. World Journal of Advanced Research and Reviews, 21(2), 1720-1736. https://doi.org/10.30574/wjarr.2024.21.2.0607
27. Taddeo, M. (2019). Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity: M. Taddeo. Minds and machines, 29(2), 187-191. https://doi.org/10.1007/s11023-019-09504-8
28. Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. Nature Machine Intelligence, 1(12), 557-560. https://doi.org/10.1038/s42256-019-0109-1
29. Tao, F., Akhtar, M. S., & Jiayuan, Z. (2021). The future of artificial intelligence in cybersecurity: A comprehensive survey. EAI Endorsed Transactions on Creative Technologies, 8(28). 10.4108/eai.7-7-2021.170285
30. Wiafe, I., Koranteng, F. N., Obeng, E. N., Assyne, N., Wiafe, A., & Gulliver, S. R. (2020). Artificial intelligence for cybersecurity: a systematic mapping of literature. IEEE Access, 8, 146598-146612. 10.1109/ACCESS.2020.3013145
31. Wirkuttis, N., & Klein, H. (2017). Artificial intelligence in cybersecurity. Cyber, Intelligence, and Security, 1(1), 103-119. https://d1wqtxts1xzle7.cloudfront.net/52464497/Artificial_Intelligence_in_Cybersecurity-libre.pdf?1491298654=&response-content-disposition=inline%3B+filename%3DArtificial_Intelligence_in_Cybersecurity.pdf&Expires=1777362362&Signature=Ptc14IJTaUEDK6SGzhucDHqjqcEYFtRYn48kDrZJBJa1i4dQkwYODwyDVWxDp8RTGxC~xbtuKTjLChUpdwq6E00nJOq-NGsNIHq1fcVpfo2I~gGH~jahU9IUanafocPc1V6HMacH5QKfpZ4JUikQWmPhRqLySOr5YY7O8-gcgTw1a9IPVFhNBHB5gWSEMbWW6XWKgZnKPO7K7WjKiBWlD9RiWbXjIYgIn8Q-TZziP57Zi4VOdEeppdg5nx7fj~TNmfIQHOlUKgkn6FZxF4nT990YvZJNp8PlqlSsIFjM4lk2qCdfpsEJpSaGxuPD27T-a6XYMjQ08olTA3EzHcN4vA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
32. Zavushchak, I. (2025). The impact of artificial intelligence on cybersecurity and data protection. Int. J. Wireless Microwave Technol.(IJWMT), 15(4), 65-72. https://orcid.org/ 0000-0002-5371-8775