TECHNOLOGIES AND ENGINEERING

Journal homepage: https://technologies-engineering.com.ua/en
Vol. 26, No. 3, 2025

Received: 07.02.2025
Revised: 23.05.2025
Accepted: 25.06.2025

UDC 004.056.5:004.8

DOI: 10.30857/2786-5371.2025.3.7

Cyber defence automation: Can AI outperform hackers?

Vitalii Yasenenko[*]
Master, Senior Software Developer
TP-Link
92618, 36 Technology Str., Irvine, USA
https://orcid.org/0009-0004-4801-9541

Abstract. Growing cyber threats in the context of digital transformation require improved methods of automating cyber defence, in particular, through the use of artificial intelligence to detect and respond to attacks. The purpose of the study was to assess the potential of artificial intelligence in the field of cyber defence automation, and to determine its effectiveness in countering advanced cyber-attacks. The study was based on methods of comparative analysis, a systematic approach, and forecasting of cyber threats using artificial intelligence. For the first time, the effectiveness of using artificial intelligence in automated cyber defence systems was comprehensively considered in the context of the current challenges of the digital environment. An analytical review of the capabilities of artificial intelligence in detecting and predicting cyber threats using neural networks, security information and event management systems, and machine learning algorithms was presented. The potential and limitations of automated response systems were determined, and the risks associated with the possibility of using artificial intelligence by attackers themselves were outlined. Special attention was paid to the analysis of the role of the human factor in interaction with automated systems – it was shown that full automation does not guarantee security without the critical participation of specialists. Based on comparative and predictive analysis, practical approaches to the balanced implementation of artificial intelligence in cyber defence systems were proposed. It was proved that artificial intelligence can be a powerful tool for automating cyber defence, but to ensure a high level of security, it is necessary to consider potential risks and constantly improve systems. The practical significance of the study was to develop recommendations for implementing effective security technologies in various cybersecurity sectors

Keywords: digital security; machine learning; engineering; media literacy; digital threat

Introduction

Since 2020, cyber threats have evolved from isolated attacks by individual hackers to highly coordinated actions by transnational cyber structures that can affect the security of entire states. The globalisation of the information space has created a favourable environment for the spread of malicious software, the organisation of cyber espionage, digital sabotage, and information and psychological operations. Cyberattacks can paralyse energy systems, transportation, financial networks, and medical institutions, creating a chain reaction that goes beyond national borders. Insufficient coordination between states in the field of digital security, lack of uniform response standards and political differences prevent effective countering threats.

Cybersecurity automation using artificial intelligence (AI) has significant potential to improve the effectiveness of protection against cyber threats. The study by O.Yu. Kotlyarov & L.L. Bortnik (2024) included a comparative analysis of modern virtual network protection systems, in particular, methodological approaches to building security mechanisms in a virtualised environment were considered. The researchers described key technologies such as virtual firewalls, network segment isolation, and intrusion detection tools in cloud infrastructure. The paper emphasised that most modern systems required not only a high level of automation, but also adaptability to the latest types of attacks. The researchers also drew attention to the importance of following the principles of multi-level protection and the need to integrate real-time monitoring tools. The conclusions emphasised that effective protection of virtual networks was possible only if technical solutions were combined with a flexible risk management methodology. These results were supported by S. Lysenko *et*

Suggested Citation:

Yasenenko, V. (2025). Cyber defence automation: Can AI outperform hackers? *Technologies and Engineering*, 26(3), 89-99. doi: 10.30857/2786-5371.2025.3.7.

[*]Corresponding author

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

al. (2024), who examined the problem of automating protection and threat detection through the integration of AI into cybersecurity infrastructure. Their paper noted that the use of automated monitoring and response systems to threats can significantly reduce the likelihood of successful attacks, but also noted the importance of human supervision for timely intervention in difficult situations.

G.G. Mykhalchenko *et al.* (2023) considered economic aspects of ensuring cybersecurity and challenges related to digitalisation. The researchers noted that effective counteraction to cyber threats required not only technical solutions, but also the development of a comprehensive security strategy that included legal and organisational measures. They stressed that only a systematic approach, which included training, implementation of security standards, and continuous monitoring, can ensure the sustainability of economic infrastructure to threats in the digital environment. In turn, the study conducted by A. Yaseen (2024) emphasised the importance of automating infrastructure management to improve cybersecurity. The researcher emphasised that automated systems that can optimise security measures not only reduce risks, but also allow quickly responding to new threats in real time. This integration of AI with infrastructure management systems is an important step towards ensuring high security reliability in organisations. In turn, the study by S. Tonhauser & J. Ristvej (2023), focused on automating processes in the context of countering cyber-attacks. They noted that automated technologies make it much easier to detect and neutralise threats, which is crucial for reducing the response time to security incidents. However, the researchers stressed the need to integrate such systems with conventional security methods to achieve maximum efficiency.

Analysis by S. Varga *et al.* (2022) supplemented previous research, emphasising that automation in cybersecurity through artificial intelligence allows security systems to adapt to new threats. However, as the researchers noted, such an application of AI should be not only technological, but also strategic, and should consider the context and specifics of organisational structures to maximise efficiency. The integration of AI into enterprise automation and cybersecurity is also the subject of research by S.K. Sundaramurthy *et al.* (2022). They focused on the fact that AI can change approaches to processing large amounts of data and providing cyber defence in the face of rapid changes. However, the researchers stressed the importance of combining technological solutions with human control to achieve an appropriate level of security. The study by M. Khan *et al.* (2024) focused on the latest advances in the application of artificial intelligence in cybersecurity, emphasising the importance of automated systems for detecting threats. The researchers pointed out that the use of AI in combination with other security tools can significantly increase the effectiveness of cyber defence.

An important addition to this topic was the study by S.S. Dasawat & S. Sharma (2023), which examined the integration of cybersecurity in startups and small businesses. The researchers stressed the importance of using AI for risk management and automation, especially in a rapidly changing cyber environment. They noted that the integration of such technologies reduced the vulnerability of businesses to attacks and contributed to more efficient scaling. Research by N.A.D. Sontan & N.S.V. Samuel (2024) highlighted the importance of integrating AI and cybersecurity, focusing on possible challenges and prospects for the development of this technology. The researchers pointed out the need for further developments to minimise the risks associated with ethical and security issues that arise when using AI in cyber defence. They also noted that, despite its significant potential, AI cannot completely replace the human factor, but must complement conventional methods of cyber defence.

Despite the growing interest in using AI in cybersecurity, most research was limited to passive data analysis and threat detection functions. Insufficient attention was paid to the role of AI as an active defender – a tool that can independently respond to attacks, block them in real time, and stay ahead of the actions of intruders. The potential of AI as an autonomous force in the fight against highly dynamic, coordinated cyber threats that go beyond national jurisdictions remains poorly understood. The purpose of the study was to investigate the capabilities of artificial intelligence as an active element of cyber defence, capable not only of detecting threats, but also of autonomously countering attacks in real time, considering the global nature of advanced cyber threats and transnational challenges in the field of digital security.

Materials and Methods

The research was mainly theoretical in nature and was aimed at systematising and generalising knowledge about the use of artificial intelligence to automate cyber defence in a global context. The chronological boundaries covered the period 2020-2025, with a particular focus on 2022-2025 publications. The study was conducted from January to April 2025. The analytical database consisted of secondary data collected from international scientific papers, reviews, preprints, and reports published in leading academic journals and repositories (Springer, Elsevier, arXiv, ResearchGate, ScienceDirect). The focus was on theoretical research and analytical reviews that highlighted the applications of machine learning, deep learning, natural language processing (NLP), generative AI, and the use of artificial intelligence in automated threat detection, risk management, proactive defence, and cyber intelligence. The study used a number of methods that allowed theoretically structuring the material and conducting an in-depth analysis. In particular, the comparative analysis was used to compare conventional and intelligent protection methods based on key criteria: effectiveness, adaptability, and accuracy of cyber threat detection. This approach highlighted the advantages and limitations of each class of methods and helped to assess their potential for use in the conditions of growing cyber threats.

To classify methods for predicting cyber threats, machine learning approaches were systematised by the following types: supervised learning, unsupported learning, and deep learning. This approach allowed theoretically substantiating the functionality of each of the approaches in accordance with the types of threats and data specifics, and determining the feasibility of their use in automated cyber defence systems. This allowed structuring the functionality of various approaches, characterising their areas of application, and determining which algorithms were most effective for specific cybersecurity tasks – from detecting anomalies to predicting potential attacks. A historical and logical method was also applied, which helped to trace the evolution of approaches to the use of AI in cyber defence, identify the main trends and patterns of industry development. An analytical and synthetic approach was used to summarise data from various sources and form a complete picture. Separately, a predictive method was applied, which identified promising areas for the development of automated cyber defence technologies that can effectively counteract new, more complex and adaptive cyber threats. The results obtained were considered from the perspective of efficiency, adaptability, and security of automated cyber defence systems, which was extremely important in the context of the constant increasing complexity of attacks and cybersecurity challenges.

Results

The use of AI to detect cyber threats: Neural networks, SIEM systems, and automated firewalls. The current landscape of cyber threats is becoming increasingly complex, and the growing number and variety of attacks require new approaches to security. In an ever-changing cyber threat environment, conventional defence methods often fail to effectively counter new attacks. That is why AI, in particular, neural networks, security information and event management (SIEM) systems, and automated firewalls, play a crucial role in detecting threats, predicting attacks, and responding quickly to incidents. These technologies not only automate security processes, but also allow systems to adapt to new, unknown threats, making them important tools in modern cybersecurity.

Neural networks have become an important tool in cybersecurity due to their ability to detect anomalies in user behaviour and network traffic. They can learn from large amounts of data and detect when certain user actions or network traffic deviate from normal patterns. This approach allows detecting both known and new types of attacks, especially those that do not yet have predefined signatures in conventional threat detection systems. Neural networks that were used to detect behavioural anomalies can adapt to new conditions and automatically adjust their threat detection algorithms, which is extremely important for dealing with new types of attacks, such as "zero-day" attacks that cannot be detected by signature detection. This technology provides higher detection accuracy and efficiency, reducing the number of false positives and improving the ability to predict attacks.

SIEM systems are the foundation for managing security information because they combine data from multiple sources, such as network devices, servers, applications, and other infrastructure components. These systems use algorithms to correlate events and identify anomalies that may indicate the presence of a threat. They allow quickly detecting security incidents and responding to them in real time, which is critical for timely protection of organisations (Sharma *et al.*, 2024). Integration of AI into SIEM systems allows automating detection and response processes and also predicting possible threats based on historical data. Machine learning algorithms can detect complex relationships between events and detect threats that may not be obvious to the human eye. For example, if some system events indicate the presence of a potential threat, the AI can predict the probability of an attack, which would allow the system to take the necessary measures at an early stage.

Automated firewalls equipped with AI provide a higher level of network protection, as they can adapt to new types of attacks and change their real-time traffic filtering strategies. Conventional firewalls use static rules to block or allow traffic, but they may not always account for new, unknown types of attacks. In turn, AI can detect traffic anomalies and take action without having to manually update security rules. Through the use of machine learning algorithms, automated firewalls can "learn" from new traffic patterns and adapt their algorithms to changes in the network environment. This allows effectively detecting new types of attacks and avoiding outdated signatures or parameters, which makes protection more dynamic and flexible. In addition, AI can use historical data about network threats to predict future attacks, thereby reducing the likelihood of successful penetration into the system. In order to compare conventional and intelligent protection methods, in particular, their effectiveness and adaptability to new threats, Table 1 was presented, which demonstrated the main differences between them.

Table 1. *Comparison of conventional and intelligent security methods*

Technology	Conventional approach	AI-based approach
Neural networks	Detection of threats using fixed signatures.	Detection of anomalies and new types of threats through analysis of user behaviour and network traffic.
SIEM-systems	Analysis of events using simple rules.	Use of machine learning algorithms to correlate events and predict attacks.
Firewalls	Static filtering rules.	Adaptive traffic filtering based on anomaly analysis and machine learning.

Source: *compiled by the author based on E. Aghaei et al. (2022)*

The integration of artificial intelligence into cyber threat detection and protection systems, such as neural networks, SIEM systems, and automated firewalls, provides more effective and adaptive protection for networks and systems. The use of AI can significantly reduce the response time to attacks, reduce the number of false positives, and identify new, unknown threats. The ability of such systems to learn and adapt makes them important tools for maintaining cybersecurity in an ever-changing threat environment. However, to achieve maximum effect, these systems must be constantly improved and adapted to new types of attacks.

Methods for predicting cyber threats based on the analysis of previous attacks using AI. In the light of the rapid evolution of cyber threats, characterised by both high dynamics and increasing complexity, cybersecurity experts face the need to move from conventional reactive protection models to proactive approaches focused on predicting potential attacks. In this context, artificial intelligence, as a set of algorithms capable of self-learning, identifying patterns and making decisions based on large-scale data analysis, becomes of strategic importance. Its use in predicting cyber threats is primarily based on the ability to study historical data on security incidents, model behavioural scenarios, and identify latent patterns inherent in the actions of attackers. As a result of the classification of machine learning methods, the features of each of the approaches were identified in the context of their effectiveness in predicting cyber threats. In particular, supervised learning proved to be suitable for classifying known types of attacks, provided that marked-up data is available, which provides high accuracy, but requires significant resources for preparing datasets. Unsupported learning effectively detects anomalies and new threats in raw data, but may have lower accuracy due to the difficulty of interpreting the results. Deep learning, due to its ability to process complex multidimensional data, has shown high efficiency in detecting hidden and complex threats, in particular, in cases of analysing network traffic or user behavioural patterns.

Analysing previous attacks using machine learning algorithms allows security systems to automatically detect patterns in the actions of cybercriminals. For example, supervised learning models are trained on labelled datasets, where each event is classified as safe or harmful. In the future, these models are used to analyse new data in real time, helping to determine the probability of a threat with high accuracy. Models that work with large sets of network traffic logs, user behavioural logs, security sensor data, etc., were particularly effective. In turn, unsupervised learning methods, in particular, clustering and dimensionality reduction, allow identifying new, previously undetected threats that have no direct analogues in training sets (Iaiani *et al.*, 2021). An important forecasting tool is time series analysis, which is used to identify repetitive patterns in the behaviour of a network or system. The use of such approaches, in particular Arima, Prophet, or recurrent neural networks (RNNs) models, allows predicting periods of increased risk, considering the seasonality of attacks or specific events (for example, large international forums, elections, releases of critical software). In this way, organisations are given the opportunity to take preventive measures in advance, strengthening protection at potentially vulnerable moments.

Special attention should be paid to the concept of threat actor profiling, which is based on the study of characteristic patterns of actions of certain hacker groups or individual attackers. Artificial intelligence analyses attack vectors, penetration methods, typical targets, and tools used to create generalised profiles of attack structures. In the future, such profiles are used to predict future actions of intruders, which allows more accurately configuring attack warning and blocking systems. Systematised forecasting methods, their functional purposes, and examples of practical application in the field of cybersecurity are given in Table 2.

Table 2. *Methods for predicting cyber threats using AI: characteristics and application examples*

Method / algorithm	Type of training	Purpose of the application	Usage examples
Supervised learning	Controlled	Classification of threats based on previous data.	Detection of phishing emails, malicious software.
Unsupervised learning	Uncontrolled	Detection of new threats and anomalies.	Detection of atypical user behaviour on the network.
Time series analysis	Depends on the model	Predicting the frequency of attacks.	Predicting Distributed Denial of Service (DDOS) attacks during peak periods.
Threat profiling (clustering)	Uncontrolled	Defining patterns of activity of hacker groups.	Building attack profiles for Advanced Persistent Threat (APT) groups.
Deep learning (RNN, LSTM)	Deep learning	Behavioural sequence analysis.	Prediction of multi-step attacks in complex systems.

Source: *compiled by the author based on S.S. Dasawat & S. Sharma (2023), R. Kaur et al. (2023)*

The integration of the above methods into integrated cyber defence systems forms a new paradigm, where the key role belongs not only to detecting ex post facto attacks, but also to timely anticipating potential threats. The use of AI in this area provides increased efficiency in detecting complex, hybrid attacks, reducing response time, and improving information security management at the strategic level. Thus, the use of artificial intelligence in predicting cyber threats not only optimises technical processes, but also strengthens the resilience of information systems to future challenges in global cyberspace.

Use of SOAR to quickly respond to cyber-attacks. Modern digital infrastructures are under constant pressure from attackers using increasingly sophisticated attack tools, including sophisticated multi-level intrusions, social engineering, zero-day exploits, and automated botnets. In such conditions, responding to cyber incidents manually loses its effectiveness, as it requires significant time, resources, and highly qualified specialists. This leads to an increasing need for the introduction of SOAR systems – technologies that combine process orchestration, automation of routine actions, and analytical support for responding to security incidents. SOAR platforms perform several key functions. First, they integrate with all elements of enterprise information security: SIEM systems, network traffic monitoring tools, antivirus programmes, cloud services, and threat intelligence sources. Thus, a single incident management point is achieved, which allows quickly responding to threats regardless of their source. Second, SOAR allows automating typical response scenarios. For example, when suspicious activity is detected, the system can automatically block an IP address or user within the corporate network; send a request for verification to the sandbox environment; notify responsible analysts through integrated communication channels; create a ticket in the incident management system and launch an investigation protocol. These actions are performed using built-in playbooks – predefined logical decision chains that can be modified to suit the specifics of the organisation (Li *et al.*, 2023).

A separate role in SOAR systems was played by AI, in particular, machine learning algorithms that are used to prioritise incidents. For example, by analysing the history of attacks, AI can determine that a certain type of traffic has already led to data leaks in the past, and give it a higher level of criticality compared to an inactive threat. AI can also identify false positives, reducing the burden on analysts. An important aspect is the ability to preserve evidence and document actions in SOAR systems. All actions are automatically logged: from the moment the threat is detected until it is completely eliminated. This creates a reliable audit chain that can be used during internal audits or as part of criminal investigations. In practice, the use of SOAR reduces the response time to incidents to several seconds or minutes; increases the level of automation of security processes by up to 80%; reduces the number of false positives due to intelligent analysis; and standardises solutions in accordance with enterprise security policies. In large corporations or government agencies (such as CERT centres), where thousands of security events are recorded daily, SOAR systems (for example, IBM QRadar SOAR, Splunk Phantom, Palo Alto Cortex XSOAR) can reduce the burden on first-level analysts by transmitting only incidents that really require human intervention. The key stages of SOAR systems' automated response to cyber incidents are shown in Figure 1.

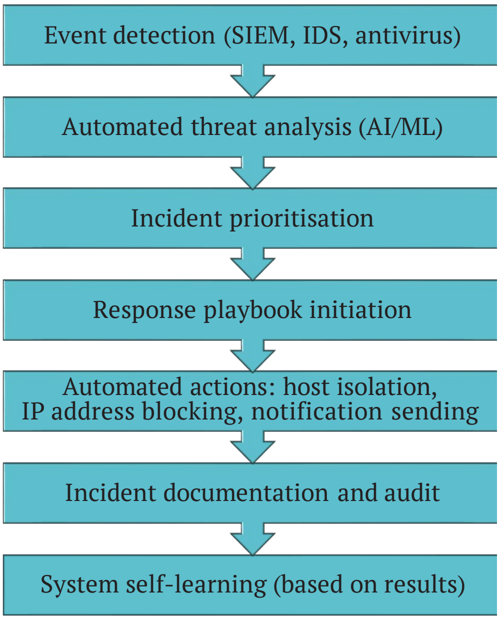
Event detection (SIEM, IDS, antivirus)
Automated threat analysis (AI/ML)
Incident prioritisation
Response playbook initiation
Automated actions: host isolation, IP address blocking, notification sending
Incident documentation and audit
System self-learning (based on results)

Figure 1. *Key stages of functioning of the SOAR system in the process of responding to cyber incidents*

Source: *compiled by the author based on I.F. Kilincer et al. (2023), R. Kaur et al. (2023)*

Thus, SOAR systems enhanced by AI transform the response to cyber-attacks from a predominantly manual process to a flexible, dynamic, and adaptive security system that can scale according to the level of threats and changes in the cyber environment. In the future, such systems will form the basis of the so-called smart cybersecurity, based on machine analysis, autonomous solutions, and a high degree of self-learning.

Cybersecurity automation risks: Ways for attackers to bypass protection using algorithms. Automation of cyber defence, in particular through the use of AI algorithms, plays a key role in strengthening the effectiveness of security systems against the background of increasing intensity of cyber threats. However, the use of automated security mechanisms introduces a number of risks, including those related to the vulnerability of the algorithms themselves, which are used for both defence and attack. This creates a technological symmetry in which attackers can use the same principles that are used to provide protection to circumvent them or undermine their effectiveness. One of the main problems is the application of adversarial machine learning, which involves targeted modification of input data, which allows attackers to mislead automated detection systems. As a result, adversarial examples are created, which, although they do not look different from normal data, actually carry a hidden threat. Such attack methods allow attackers to manipulate the results of AI algorithms, bypassing security systems, which significantly reduces the effectiveness of conventional methods of detecting threats. They distort behavioural signals, making them impossible to properly interpret within anomaly detection systems.

An important aspect is malware self-learning, which is based on machine learning algorithms that can adapt to changes in the environment and evolve in response to new challenges. In particular, polymorphic viruses can change their signatures as they spread, which allows them to avoid detection by antivirus programmes based on signature analysis. This adaptability makes it difficult to detect malware and significantly increases their ability to avoid conventional cyber defences. In addition, the use of behavioural spoofing allows malware to simulate normal activity, which further complicates the operation of systems based on the analysis of behavioural anomalies. The danger of automated systems also lies in the imperfection of autonomous decisions that are made without human intervention. Despite the potential of automation to ensure prompt response, such systems can lead to false positives that have disastrous consequences for the organisational infrastructure. Excessive aggressiveness of automated mechanisms can lead to blocking legitimate user activity that disrupts the normal activities of organisations, or to missing new threats that were not considered in the system training process. The problem is that many automated systems are unable to adapt to new, unforeseen scenarios, making them vulnerable to evolutionary threats (Alamro *et al.*, 2023).

In addition, algorithms used in cybersecurity systems can become targets of attacks at the training stage, including data poisoning. Attackers, knowing the structure of models, can intentionally enter corrupt or erroneous data at the training stage, which leads to a violation of their ability to correctly interpret information in a real environment. This can have negative consequences for the accuracy of systems, especially at the stages of detecting new or unknown threats, when erroneous training leads to incorrect response to real cyber threats. Table 3 provided a brief overview of the key risks associated with cybersecurity automation, and potential ways to reduce them.

Table 3. *Main risks of cyber defence automation and ways to minimise them*

Type of risk	Essence of the problem	Possible counteraction measures
Reverse machine learning	Deception of AI algorithms using specially prepared input data.	Use of secure models, anomaly detection, regular testing.
Polymorphic malware	Automatic signature change to avoid detection.	Behavioural analysis, combining signature and heuristic approaches.
Excessive system autonomy	Erroneous actions of an automated system without human intervention.	Implementation of manual confirmation for critical actions.
Process disruption due to an error	Automatic blocking of legitimate activity due to false positives.	Multi-level monitoring, hybrid response scenarios.
Data poisoning	Impact on the learning process to distort future decisions.	Verification of data sources, filtering abnormal samples, and monitoring models.

Source: *compiled by the author based on L.F. Sikos (2023)*

As a result, although automation in cyber defence offers significant progress in the speed and efficiency of incident response, its use is accompanied by numerous risks that require careful monitoring and continuous improvement of protection mechanisms. The use of artificial intelligence, combined with flexible security approaches that include the human factor, remains critical to ensuring a reliable and sustainable response to modern cyber threats.

Functions of artificial intelligence in cybersecurity. In the context of the rapid evolution of cyber threats and high requirements for the effectiveness of information system protection, the role of AI in the field of cybersecurity is becoming increasingly important. Advanced technologies allow artificial intelligence to perform functions that previously required significant human resources, and have significant potential to improve the effectiveness of security systems. However, there is a question whether AI can completely replace the human factor, or whether it remains only an auxiliary tool at the disposal of cybersecurity specialists. In this context, it is important to consider the role of AI in automating cybersecurity, its potential to replace humans in threat detection and response processes, and the limitations and challenges that arise from its use Artificial intelligence offers a number of opportunities to automate routine and complex cybersecurity processes, significantly reducing the need for human intervention at the stages of detecting and responding to threats. Machine learning-based systems, such as neural networks, can automatically detect anomalies in user behaviour and network traffic by analysing huge amounts of data in real time. They are able to detect potential threats much faster than humans, which is a critical factor for timely neutralisation of attacks, especially in the case of DDoS or APT attacks. The use of SIEM systems with AI integration allows automating the processes of collecting, analysing, and correlating data from various security sources (network devices, servers, databases). With the help of such technologies, AI can not only detect familiar threats, but also identify new attack patterns that may indicate complex intrusion methods used by attackers. This allows security systems to respond in real time, which is directly important for preventing or minimising damage from attacks. Based on the ability to automatically block suspicious actions, such as unauthorised access attempts or configuration changes, AI-based systems can quickly take countermeasures without human intervention. This allows significantly reducing the response time to incidents, which is important for successful protection of critical infrastructures, where every minute matters (Sharma *et al.*, 2024).

However, despite the power of AI, full automation of cybersecurity is unlikely. AI, even with all its advances in machine learning and data analysis, has significant limitations, especially when it comes to new, unknown, or extremely complex threats. For example, detecting APT attacks or complex phishing attacks using multiphase social engineering techniques requires deep strategic thinking and contextual analysis, which is not always available to automated systems. Cybersecurity professionals can interpret signals coming from systems based on their experience and knowledge of new threats that are not obvious to algorithms. They can consider the context of the event, its priority, and interaction with other incidents, which allows them to make more informed decisions about further actions. In addition, human expertise is needed to adapt and customise AI models, as algorithms require regular training on new data and adjustments to meet changes in the threat environment. It should also be noted that the ethical and legal aspects of using AI in cybersecurity can be important to ensure compliance with national and international security and privacy standards. The human factor is important for developing risk management strategies and ensuring compliance with legal requirements (Kaur *et al.*, 2023).

The most effective approach to cybersecurity is to combine the capabilities of artificial intelligence and human expertise. AI can be used to automate routine tasks such as network monitoring, detecting standard threats, analysing huge amounts of data, and responding to threats in real time. This frees cybersecurity professionals from the need to perform low-level operations, which increases their effectiveness and allows them to focus on more complex tasks, such as developing security strategies, evaluating new vulnerabilities, or analysing new types of attacks. In particular, due to the use of AI, it is possible to significantly reduce the response time to incidents, increase the accuracy of threat detection, and provide more flexible adaptation of security systems to new conditions. However, decisions that require deep strategic analysis should remain in the hands of specialists who can assess the interaction of attacks with other elements of the system, and anticipate possible consequences for the organisation as a whole. Artificial intelligence is an essential tool in modern cybersecurity, allowing automating a number of critical processes, such as threat detection, incident response, and data analysis. However, it cannot completely replace the human factor, because successful detection of complex, unknown, or multiphase attacks requires human intuition, strategic thinking, and the ability to assess the context of events. Therefore, the best approach is to integrate AI into the cybersecurity system, where it acts as a powerful auxiliary tool for specialists, reducing their workload and ensuring a faster and more effective response to threats.

Discussion

An analysis of the literature on the use of AI in cybersecurity has demonstrated various approaches to solving problems of protecting information systems and infrastructures. The study conducted by A. Mohammed (2023) examined the AI paradox in cybersecurity, emphasising that artificial intelligence can act as both a powerful defender and a potential tool for exploiting cyber threats. The researcher focused on the fact that although AI can improve the response to attacks due to the ability to quickly analyse large amounts of data, attackers can also use these technologies to bypass security systems. These results are consistent with current results that highlight the importance of balancing the use of technologies for protection and the risks associated with their possible abuse.

The study by N.P.O. Shoetan *et al.* (2024) presented a conceptual model of the impact of AI on cybersecurity in the telecommunications sector. The researchers stressed that the introduction of AI significantly improves the ability of systems to detect threats, predict attacks, and reduce response time. These results are consistent with data from the current study, which also points to the positive effect of automation and intelligence in the field of cyber defence. The current study looked more broadly at automation as a tool not only for telecommunications security, but also for critical infrastructure in general. W.S. Admass *et al.* (2023) provided an overview of the current state of cybersecurity, the main challenges and areas of its development. The researchers highlighted the complexity of attacks, the lack of qualified personnel, and the limitations of AI adaptation. These findings are partially consistent with current results, especially with regard to the problem of personnel and the importance of automation as a compensatory mechanism in the face of a shortage of specialists. The study added to this the emphasis on reducing the human factor through the introduction of automated tools, which allowed increasing the effectiveness of real-time protection.

V. Švábenský *et al.* (2023) investigated automated feedback in cybersecurity training programmes. Although the researchers focused on the educational field, their findings coincided with the results of the present study, focusing on the use of automated systems to improve the quality of response to incidents. The research reflected a similar idea in the context of production and corporate environments, where training and automation of threat detection processes were critical. S. Vyas *et al.* (2023) proposed a generalised review of automated cyber defence systems, in particular, they analysed technologies that provide early detection of attacks and automatic response. Their findings were fully consistent with the current study, especially in terms of the role of AI in reducing the burden on security analysts, improving response speed, and reducing error rates. N.A. Folorunso *et al.* (2024) focused on the impact of AI on security compliance. They noted that AI-based systems can help not only detect violations, but also ensure compliance with regulations through constant monitoring. This provision correlates with current results, which emphasised that automated systems can maintain a high level of compliance with safety standards through continuous monitoring and flexible response. In this context, S. Ahmadov (2024) emphasised that encryption remains a core tool for securing personal data in cloud environments. The study highlighted the role of effective key management and compliance with data protection laws, reinforcing the importance of a multi-layered approach to cybersecurity.

The study by J. Xu *et al.* (2024) introduced Autoattacker system, which used the capabilities of large language models to automate cyber-attacks. The researchers demonstrated the potential of the Large Language Models (LLM) not only as a security tool, but also as a threat that requires a new level of caution in implementing AI in cybersecurity systems. These findings were consistent with the current study, which also highlighted the twofold nature of AI use – both for protection and as a potential threat in the hands of attackers. This study, however, focused more on the protective aspect and suggested approaches to minimising risks. In turn, the review by J.P. Bharadiya (2023) focused on the future of cybersecurity in the context of Web 3.0, emphasising the role of machine learning in shaping safer digital environments. The researcher noted that machine learning can increase the adaptability of cyber defence systems. The present results also indicated the ability of machine learning systems to quickly learn from new threats and adapt real-time protection methods, which confirmed the relevance of the approaches proposed by J.P. Bharadiya.

N.U. Prince *et al.* (2024) investigated data-driven methods in AI cybersecurity that can improve the accuracy and speed of attack detection. Their emphasised on processing large amounts of data and using predictive analytics directly echoed current approaches, which also emphasised the role of large-scale data processing in creating adaptive protection. F. Mahmud *et al.* (2025) examined the use of AI in IT project management with a focus on threat detection and risk reduction. Their findings demonstrated the effectiveness of AI integration at the planning and risk assessment stages, which was also important for the current study, as it also demonstrated that proactive AI-based risk assessment can reduce the impact of incidents on critical systems. A. Shahana *et al.* (2024) in their paper focused on the balance between implementing AI in cybersecurity and providing safeguards to avoid abuse and technical vulnerabilities. The researchers emphasised the risks of full automation of protection and the need for human control. Similar to the current results, the researchers recognised that an effective system should be hybrid – a combination of AI and the human factor. It was also noted that excessive automation without ethical and legal barriers can lead to threats. In turn, N.A.O. Adewusi *et al.* (2022) considered the specific use of AI for cyber defence in the agricultural sector – “smart farms”. AI has been shown to automatically detect threats in Internet of Things (IoT) infrastructures. This study had a broader application, but it also focused on the vulnerability of IoT systems. N.A.O. Adewusi *et al.* supplemented the current study with examples from a specific sector. M. Rizvi (2023) demonstrated the benefits of AI in real-time threat detection, especially through continuous learning and behaviour pattern analysis. The current study also highlighted the adaptability of AI systems that can identify anomalies even before they escalate into full-fledged attacks. The findings of M. Rizvi were fully consistent with these observations.

A.B. Pandey *et al.* (2022) conducted a comprehensive review of general cybersecurity trends, including AI applications, new attack vectors, and security concepts. The current study also highlighted certain trends: the development of threats, the role of artificial intelligence, and a paradigm shift in security. Y. Yigit *et al.* (2024) considered the use of generative AI in cybersecurity – both in defensive and attacking aspects (deepfake, deception of recognition systems, etc.). This was a significant addition to the analysis, which considered AI in a classical context, but generative models created a new level of challenges that deserves further inclusion in this discussion. M. Elsisi & M. Tran (2021) described an IoT architecture that uses deep neural networks to protect automated transport systems from cyber-attacks. The research of scientists is more general, but at the same time it demonstrates the successful application of DNN in real-world cyber-physical systems, which confirms current assumptions about the effectiveness of deep learning in defensive scenarios.

Thus, comparison with other papers showed that the study is consistent with global trends in the investigation of the role of AI in the automation of cyber defence, while contributing to the understanding of how these approaches can be adapted to different sectors and infrastructures. Unlike many reviewed works, it places stronger emphasis on the integration of adaptive mechanisms and blockchain technologies to enhance transparency and resilience. The study also highlights the importance of balancing automation with human oversight, which is crucial for maintaining ethical and secure AI deployment in cybersecurity.

Conclusions

As part of the study of the use of neural networks, SIEM systems, and automated firewalls, it was found that the integration of deep learning significantly increases the accuracy of detecting complex and unknown threats, but simultaneously creates new challenges for controlling algorithms. In the section on methods for predicting cyber threats, it was found out that the most effective systems are those that use historical attack patterns in combination with adaptive algorithms – such approaches allow predicting potential threat vectors with high reliability. SOAR analysis has demonstrated their ability to reduce incident response time to a few seconds, but effectiveness remains dependent on the quality of the initial data and the correct configuration of scenarios. The section on automation risks showed that attackers are already experimenting with using generative models to bypass security and create adaptive malware that requires constant updating of security systems. Ultimately, in the debate about whether AI can replace the human factor, it is concluded that at the moment it is not a replacement, but a symbiosis: artificial intelligence is effective in repetitive tasks and processing large amounts of data, but critical thinking, ethics, and strategic decisions are left to humans. Recommendations for improving the effectiveness of automated cyber defence systems consist in combining AI technologies with the human factor, which will increase the accuracy of threat detection and the effectiveness of responding to them. The key is to improve prediction algorithms to more accurately identify new types of attacks, and develop systems that can detect and neutralise complex threats, including those that use AI to bypass existing defences. Attention should be paid to the ethical and legal aspects of using AI in cybersecurity, which would ensure the safe and responsible use of technologies. Areas for further research include the development of autonomous security systems that can respond independently to new threats, and the creation of security models for new technologies such as IoT and 5G. Investigation of the interaction of AI with other innovative technologies, in particular, blockchain and quantum computing, can lead to more sustainable security systems. It is also important to study adaptive systems that can dynamically adjust their strategies to suit new types of attacks and analyse hybrid attacks that combine cybernetic and physical elements.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Adewusi, N.A.O., Chiekezie, N.N.R., & Eyo-Udo, N.N.L. (2022). The role of AI in enhancing cybersecurity for smart farms. *World Journal of Advanced Research and Reviews*, 15(3), 501-512. doi: 10.30574/wjarr.2022.15.3.0889.

[2] Admass, W.S., Munaye, Y.Y., & Diro, A.A. (2023). Cyber security: State of the art, challenges and future directions. *Cyber Security and Applications*, 2, article number 100031. doi: 10.1016/j.csa.2023.100031.

[3] Aghaei, E., Niu, X., Shadid, W., & Al-Shaer, E. (2022). Securebert: A domain-specific language model for cybersecurity. In F. Li, K. Liang, Z. Lin & S.K. Katsikas (Eds.), *Proceedings of the 18th EAI international conference "Security and privacy in communication networks"* (pp. 39-56). Cham: Springer. doi: 10.1007/978-3-031-25538-0_3.

[4] Ahmadov, S. (2024). Data encryption as a method of protecting personal data in a cloud environment. *Bulletin of Cherkasy State Technological University*, 29(3), 31-41. doi: 10.62660/bcstu/3.2024.31.

[5] Alamro, H., Mtouaa, W., Aljameel, S., Salama, A.S., Hamza, M.A., & Othman, A.Y. (2023). Automated Android malware detection using optimal ensemble learning approach for cybersecurity. *IEEE Access*, 11, 72509-72517. doi: 10.1109/access.2023.3294263.

[6] Bharadiya, J.P. (2023). AI-driven security: How machine learning will shape the future of cybersecurity and web 3.0. *American Journal of Neural Networks and Applications*, 9(1), 1-7. doi: 10.11648/j.ajnna.20230901.11.

[7] Dasawat, S.S., & Sharma, S. (2023). Cyber security integration with smart new age sustainable startup business, risk management, automation and scaling system for entrepreneurs: An artificial intelligence approach. In *2023 7th international conference on intelligent computing and control systems (ICICCS)* (pp. 1357-1363). Madurai: Institute of Electrical and Electronics Engineers. doi: 10.1109/ICICCS56967.2023.10142779.

[8] Elsisi, M., & Tran, M. (2021). Development of an IoT architecture based on a deep neural network against cyber attacks for automated guided vehicles. *Sensors*, 21(24), article number 8467. doi: 10.3390/s21248467.

[9] Folorunso, N.A., Adewumi, N.T., Adewa, N.A., Okonkwo, N.R., & Olawumi, N.T.N. (2024). Impact of AI on cybersecurity and security compliance. *Global Journal of Engineering and Technology Advances*, 21(1), 167-184. doi: 10.30574/gjeta.2024.21.1.0193.

[10] Iaiani, M., Tugnoli, A., Bonvicini, S., & Cozzani, V. (2021). Analysis of cybersecurity-related incidents in the process industry. *Reliability Engineering & System Safety*, 209, article number 107485. doi: 10.1016/j.ress.2021.107485.

[11] Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, 97, article number 101804. doi: 10.1016/j.inffus.2023.101804.

[12] Khan, M.I., Arif, A., & Khan, A.R.A. (2024). The most recent advances and uses of AI in cybersecurity. *BULLET: Journal of Multidisciplinary Science*, 3(4), 566-578.

[13] Kilincer, I.F., Ertam, F., Sengur, A., Tan, R., & Acharya, U.R. (2023). Automated detection of cybersecurity attacks in healthcare systems with recursive feature elimination and multilayer perceptron optimization. *Journal of Applied Biomedicine*, 43(1), 30-41. doi: 10.1016/j.bbe.2022.11.005.

[14] Kotlyarov, O.Yu., & Bortnik, L.L. (2024). Comparative analysis of modern virtual network protection systems and their methodologies. *Modern Information Security*, 4(60), 60-72. doi: 10.31673/2409-7292.2024.040007.

[15] Li, G., Ren, L., Fu, Y., Yang, Z., Adetola, V., Wen, J., Zhu, Q., Wu, T., Candan, K., & O'Neill, Z. (2023). A critical review of cyber-physical security for building automation systems. *Annual Reviews in Control*, 55, 237-254. doi: 10.1016/j.arcontrol.2023.02.004.

[16] 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. doi: 10.46852/0424-2513.1.2024.6.

[17] Mahmud, F., Barikdar, C.R., Hassan, J., Goffer, M.A., Das, N., Orthi, S.M., Kaur, J., Hasan, S.N., & Hasan, R. (2025). AI-driven cybersecurity in IT project management: Enhancing threat detection and risk mitigation. *Journal of Posthumanism*, 5(4), 23-44. doi: 10.63332/joph.v5i4.974.

[18] Mohammed, A. (2023). The paradox of AI in cybersecurity: Protector and potential exploiter. *Baltic Journal of Engineering and Technology*, 2(1), 70-76.

[19] Mykhalchenko, G.G., Snitko, Y.M., & Ivanenko, V.O. (2023). Cyber security in the economy: Protection against cyber threats in a digitalized world. *Scientific Notes of Lviv University of Business and Law*, 38, 377-384.

[20] Pandey, A.B., Tripathi, A., & Vashist, P.C. (2022). A survey of cyber security trends, emerging technologies and threats. In R. Agrawal, J. He, E. Shubhakar Pilli & S. Kumar (Eds.), *Cyber security in intelligent computing and communications* (pp. 19-33). Singapore: Springer. doi: 10.1007/978-981-16-8012-0_2.

[21] Prince, N.U., Faheem, M.A., Khan, O.U., Hossain, K., Alkhayyat, A., Hamdache, A., & Elmouki, I. (2024). AI-powered data-driven cybersecurity techniques: Boosting threat identification and reaction. *Nanotechnology Perceptions*, 20(S10), 332-353. doi: 10.62441/nano-ntp.v20is10.25.

[22] Rizvi, M. (2023). Enhancing cybersecurity: The power of artificial intelligence in threat detection and prevention. *International Journal of Advanced Engineering Research and Science*, 10(5), 55-60. doi: 10.22161/ijaers.105.8.

[23] Shahana, A., Hasan, R., Farabi, S.F., Akter, J., Mahmud, M.A.A., Johora, F.T., & Suzer, G. (2024). AI-driven cybersecurity: Balancing advancements and safeguards. *Journal of Computer Science and Technology Studies*, 6(2), 76-85. doi: 10.32996/jcsts.2024.6.2.9.

[24] Sharma, S., Dutta, N., & Vegesna, V.V. (2024). *Examining ChatGPT's and other models' potential to improve the security environment using generative AI for cybersecurity*. doi: 10.15680/IJMRSET.2024.0709031.

[25] Shoetan, N.P.O., Amoo, N.O.O., Okafor, N.E.S., & Olorunfemi, N.O.L. (2024). Synthesizing AI's impact on cybersecurity in telecommunications: A conceptual framework. *Computer Science & IT Research Journal*, 5(3), 594-605. doi: 10.51594/csitrj.v5i3.908.

[26] Sikos, L.F. (2023). Cybersecurity knowledge graphs. *Knowledge and Information Systems*, 65, 3511-3531. doi: 10.1007/s10115-023-01860-3.

[27] Sontan, N.A.D., & Samuel, N.S.V. (2024). The intersection of artificial intelligence and cybersecurity: Challenges and opportunities. *World Journal of Advanced Research and Reviews*, 21(2), 1720-1736. doi: 10.30574/wjarr.2024.21.2.0607.

[28] Sundaramurthy, S.K., Ravichandran, N., Inaganti, A.C., & Muppalaneni, R. (2022). The future of enterprise automation: Integrating AI in cybersecurity, cloud operations, and workforce analytics. *Artificial Intelligence and Machine Learning Review*, 3(2), 1-15.

[29] Švábenský, V., Vykopal, J., Čeleda, P., & Dovjak, J. (2023). Automated feedback for participants of hands-on cybersecurity training. *Education and Information Technologies*, 29, 11555-11584. doi: 10.1007/s10639-023-12265-8.

[30] Tonhauser, M., & Ristvej, J. (2023). Cybersecurity automation in countering cyberattacks. *Transportation Research Procedia*, 74, 1360-1365. doi: 10.1016/j.trpro.2023.11.283.

[31] Varga, S., Sommestad, T., & Brynielsson, J. (2022). Automation of cybersecurity work. In T. Sipola, T. Kokkonen & M. Karjalainen (Eds.), *Artificial intelligence and cybersecurity: Theory and applications* (pp. 67-101). Cham: Springer. doi: 10.1007/978-3-031-15030-2_4.

[32] Vyas, S., Hannay, J., Bolton, A., & Burnap, P.P. (2023). *Automated cyber defence: A review*. doi: 10.48550/arXiv.2303.04926.

[33] Xu, J., Stokes, J.W., McDonald, G., Bai, X., Marshall, D., Wang, S., Swaminathan, A., & Li, Z. (2024). *AutoAttacker: A large language model guided system to implement automatic cyber-attacks*. doi: 10.48550/arXiv.2403.01038.

[34] Yaseen, A. (2024). Enhancing cybersecurity through automated infrastructure management: A comprehensive study on optimizing security measures. *Quarterly Journal of Emerging Technologies and Innovations*, 9(1), 38-60.

[35] Yigit, Y., Buchanan, W.J., Tehrani, M.G., & Maglaras, L. (2024). *Review of generative AI methods in cybersecurity*. doi: 10.48550/arXiv.2403.08701.

Автоматизація кіберзахисту: чи може ШІ випередити хакерів?

Віталій Ясененко

Магістр, старший розробник програмного забезпечення
TP-Link
92618, вул. Технологій, 36, м. Ірвайн, США
https://orcid.org/0009-0004-4801-9541

Анотація. Зростаючі кіберзагрози в умовах цифрової трансформації вимагають вдосконалення методів автоматизації кіберзахисту, зокрема через використання штучного інтелекту для виявлення і реагування на атаки. Метою дослідження була оцінка потенціалу штучного інтелекту у сфері автоматизації кіберзахисту, а також визначення його ефективності в протидії сучасним кібератакам. Дослідження базувалось на методах порівняльного аналізу, системного підходу та прогнозування кіберзагроз за допомогою штучного інтелекту. Вперше комплексно розглянуто ефективність застосування штучного інтелекту в автоматизованих системах кіберзахисту з урахуванням поточних викликів цифрового середовища. Представлено аналітичний огляд можливостей штучного інтелекту у виявленні та прогнозуванні кіберзагроз за допомогою нейромереж, систем управління інформацією та подіями безпеки і алгоритмів машинного навчання. Визначено потенціал і обмеження систем автоматизованого реагування, а також окреслено ризики, пов'язані з можливістю використання штучного інтелекту самими зловмисниками. Особливу увагу приділено аналізу ролі людського фактора у взаємодії з автоматизованими системами – показано, що повна автоматизація не гарантує безпеку без критичної участі спеціалістів. На основі порівняльного та прогностичного аналізу запропоновано практичні підходи до збалансованого впровадження штучного інтелекту в системи кіберзахисту. Доведено, що штучний інтелект може бути потужним інструментом для автоматизації кіберзахисту, проте для забезпечення високого рівня безпеки необхідно враховувати потенційні ризики та постійно вдосконалювати системи. Практичне значення дослідження полягає в розробці рекомендацій для впровадження ефективних технологій захисту в різних секторах кібербезпеки

Ключові слова: цифрова безпека; машинне навчання; інжиніринг; медіаграмотність; цифрова загроза