UDC 004.8 **DOI:10.30837/rt.2025.1.220.03**

Yu. L. GOLIKOV

STUDY OF THE CURRENT STATE AND PROSPECTS OF ARTIFICIAL INTELLIGENCE IN CYBERSECURITY

Introduction

Artificial intelligence (AI) has been changing the cybersecurity space for more than a decade, thanks to machine learning (ML), which accelerates threat detection and detects anomalous behavior of users and objects. That's why AI is gradually becoming an integral part of modern cybersecurity systems, helping to identify threats, automate processes, and increase the level of information protection.

One of the important challenges in using AI for cybersecurity is building trust. The data used to train AI/ML models drives the output of the models. If the training data does not reflect the "real world," the model may distort its ability to deliver the expected results. Some data, such as threat information, "good" and "bad" file characteristics, compromise indicators, etc., are for everyone to see.

Another major challenge is data security. It is important to define and control what training data can be shared and what data remains secret within organizations. In the wrong hands, this data can help attackers in their attacks to undermine the ability of AI/ML to identify their files, programs, and behavior as invalid. In this regard, governments and businesses need to develop regulations, standards, and best practices to prevent new AI threats.

For example, NIST [17, 18] is already leading and participating in the development of technical standards, including international standards, that promote innovation and public trust in systems that use AI. A wide range of standards for AI data, performance, and governance is - and increasingly will be - a priority for reliable and responsible AI.

NIST has developed a plan for global engagement with the promotion and development of AI standards. The goal is to stimulate the development and implementation of consensus standards related to artificial intelligence, collaboration and coordination, and information sharing. Reflecting input from the public and private sectors. On July 26, 2024, after reviewing public comments on the project plan, NIST published the Global AI Standards Engagement Plan [17].

The development of AI technologies has brought not only new opportunities, but also new risks and dangers. The community of scientists and researchers around the world is concerned and warns of potential problems with the spread of AI. In particular, a recent report by the UK National Cyber Security Center (NCSC) [14] warns that over the next two years, AI technologies will likely increase the dynamics of cyberattacks and increase their impact on existing cryptosystems and information security tools. According to the Center [14], AI opens up great opportunities in this area even for those cryptanalysts who do not have the appropriate technical skills. And they argue that after 2025 and beyond, when AI has been trained successfully enough, AI will almost certainly improve, providing faster and more accurate cyber operations.

The role of AI will only grow in the coming years. For example, AI is able to detect threats in real time, meaning it can analyze huge amounts of data, identifying anomalous activities and potential threats faster than humans can. AI algorithms can not only identify threats but also instantly block malicious traffic or change network access rules. What is also important, automated solutions help to avoid mistakes that can occur due to human inattention or lack of skills.

Therefore, the purpose of this article is to study the current state of AI in cybersecurity, as well as the threats and risks associated with its use. The article discusses modern anti-virus solutions, the most popular AI attacks and methods of protodiagnostics. It also discusses the prospects of using AI and proves that in the modern world, AI is an integral part of cybersecurity.

1. Main areas of AI application in cybersecurity

Below, we will consider six main possible options for using AI in the field of cybersecurity (Fig. 1):

1. Automated threat detection and response. Traditional cybersecurity systems require constant monitoring and analysis of large amounts of data, which is difficult to implement manually. AI can perform the following tasks:

- Detect anomalous activity in the network using machine learning algorithms.
- Analyze user and device behavior to detect suspicious activity.
- Automatically respond to threats by blocking potentially dangerous actions without human intervention.

An example of such systems is, for example, an AI-enabled SIEM (Security Information and Event Management) system that analyzes events in real time and warns of possible attacks. SIEM systems will be discussed in more detail in Section 3 of this article.

2. Use of AI in malware analysis. Traditional antiviruses use signatures to detect malware, which makes them less effective against new attacks. Instead, AI allows:

- Use behavioral analysis to detect new viruses and Trojans.
- Recognize different types of attacks that change their code to bypass antiviruses.
- Automatically create and update threat databases without the need for manual changes.

For example, AI antiviruses such as Microsoft Defender ATP or Darktrace are gaining popularity, analyzing program behavior and detecting threats without human intervention. We discuss these methods in more detail in Section 4 of this article.

3. Strengthening cryptography and post-quantum security. It is well known that with the development of quantum computers, traditional cryptographic algorithms can become vulnerable. AI, in turn, can help in:

- Creating adaptive cryptographic solutions that can change keys and algorithms in real time.
- Optimization of encryption and decryption, which will make cryptographic systems more efficient.
- Development of new post-quantum cryptographic algorithms that will be resistant to attacks by quantum computers.

Therefore, AI is actively used to test new strong ciphers. Encryption and key management optimization – machine learning algorithms can improve the efficiency of cryptographic protocols, providing faster and more secure encryption. For example, NIST is actively researching encryption and signature algorithms for post-quantum cryptography [1], and AI can accelerate their testing and implementation [18].

4. AI in countering phishing attacks. Phishing is one of the most widespread cyber threats that is becoming increasingly complex. AI can help in:

- Automatically recognize suspicious emails and URLs.
- Analyzing the language and structure of messages to detect fraudulent schemes.
- Improved authentication mechanisms, for example, through behavioral biometrics or dynamic captchas.

One of the most well-known examples of AI in the fight against phishing is Google, which uses AI in Gmail to filter phishing emails, which allows it to block more than 99% of malicious messages.

5. Using AI to predict attacks. Artificial intelligence can help not only respond to attacks but also predict them by analyzing huge amounts of data:

- Using big data analysis to identify trends in cyberattacks.
- Creating predictive models that allow us to anticipate potential system vulnerabilities.
- Automatic analysis of the Dark Web to identify new threats and hacking tools.

For example, IBM Watson for Cyber Security [7] uses cognitive analysis to identify new threats by analyzing millions of articles, forums, and messages on the darknet.

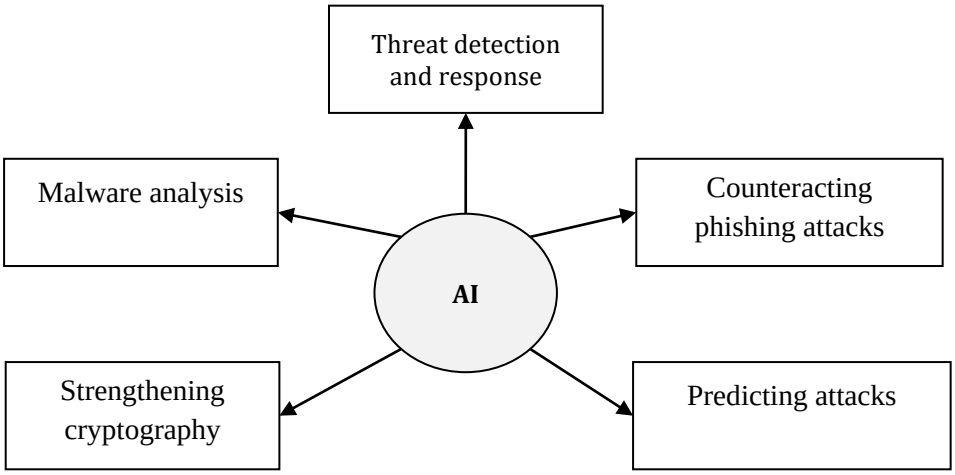

Fig. 1. The use of AI in cybersecurity

2. AI-powered protection and improved cybersecurity efficiency

Artificial intelligence (AI)-based intrusion detection systems (IDSs) are modern cybersecurity tools that use machine learning and data analytics to identify and prevent unauthorized activities on networks and systems. They are able to adapt to new threats by analyzing large amounts of data and detecting anomalies that may indicate potential attacks.

There are two main methods used in IDS:

1) Detection based on known algorithms. This method involves comparing incoming traffic with a database of known attacks. If a match is found, the system generates an alert. However, this approach is limited in detecting new or modified attacks that do not yet have corresponding algorithms.

2) Anomaly-based detection: This method uses models of normal system behavior. Deviations from this norm are considered as potential threats. AI plays a key role in building and updating such models, which allows detecting even previously unknown attacks.

It is clear that the integration of artificial intelligence into intrusion detection systems significantly increases the effectiveness of cybersecurity, allowing for the detection and prevention of both known and emerging threats. However, in order to maximize the potential of such systems, it is necessary to take into account possible limitations and ethical aspects of their application, since, for example, the use of AI may raise privacy issues, especially if the systems analyze personal data without proper control.

3. SIEM system with AI support

Security Information and Event Management, or SIEM, is a security solution that helps organizations recognize and address potential security threats and vulnerabilities before they have a chance to disrupt business operations [7].

SIEM systems help enterprise security teams detect anomalies in user behavior and use artificial intelligence (AI) to automate many of the manual processes involved in threat detection and incident response. Currently, the most well-known modern SIEM systems are the following software: Splunk Enterprise Security (Splunk), Elastic Security, IBM QRadar SIEM, Wazuh SIEM, Microsoft Sentinel.

The initial SIEM platforms were log management tools. They combined the functions of security information management (SIM) and security event management (SEM). These platforms provided real-time monitoring and analysis of security-related events. The term SIEM was coined in 2005 to describe the combination of SIM and SEM technologies.

They also made it easier to track and record security data for compliance or audit purposes. Over the years, SIEM software has evolved to include user and object behavior analytics, as well as other advanced security analytics, artificial intelligence, and machine learning capabilities to detect

anomalous behavior and indicators of advanced threats. Today, SIEM has become a core element of modern security operations centers (SOCs) for security monitoring and compliance management.

The stages of SIEM systems (Fig. 2) can be divided as follows [6]:

- Data collection – All sources of network security information, such as servers, operating systems, firewalls, antivirus software, and intrusion prevention systems, are configured to send event data to the SIEM tool. Most modern SIEM tools use agents to collect event logs from enterprise systems, which are then processed, filtered, and sent to the SIEM. Some SIEMs allow you to collect data without agents. For example, Splunk [8] offers agentless data collection on Windows using WMI.

AI SIEM systems start by aggregating data from various sources, such as network devices, servers, databases, and applications. Once ingested, the raw data is converted into a standardized format to ensure consistent and accurate data analysis regardless of the source. AI and ML significantly automate these processes, increasing the speed and intelligence with which security data is aggregated and normalized, reducing manual effort and time [9].

- Policies – The SIEM administrator creates a profile that defines the behavior of enterprise systems under normal conditions and during predefined security incidents. SIEMs provide standard rules, alerts, reports, and dashboards that can be customized to meet specific security needs.
- Data Consolidation and Correlation – SIEM solutions consolidate, analyze, and parses log files. Events are then categorized based on the raw data and correlation rules are applied to combine individual data events into meaningful security issues.
- AI SIEM systems use predictive analytics to forecast potential future threats by analyzing historical security data and identifying patterns. This capability allows organizations to proactively protect their systems rather than reacting to threats as soon as they occur. This knowledge base allows the artificial intelligence models that are at the heart of the solution to create increasingly accurate security responses and incident prevention approaches as time passes and new data is collected.

Continuously learning from past problems improves the accuracy and reliability of AI-based SIEM systems against increasingly powerful cyber threats. Ultimately, an AI-powered SIEM integrates various components such as AI, ML, deep learning, NLP, and UEBA that extend the capabilities of a traditional SIEM. This integration leads to smarter, more efficient and proactive cybersecurity measures, which is crucial in the ever-changing cyber threat environment [9].

- Alerts – If an event or set of events triggers a SIEM rule, the system notifies security personnel. When a threat is detected, AI enables SIEM systems to automate parts of the incident response process. This includes automatically triggering alerts, implementing predefined response actions, or organizing complex response workflows. One such example is an automated dynamic workflow, where the workflow that is established after a potential threat is tailored to the threat in question.

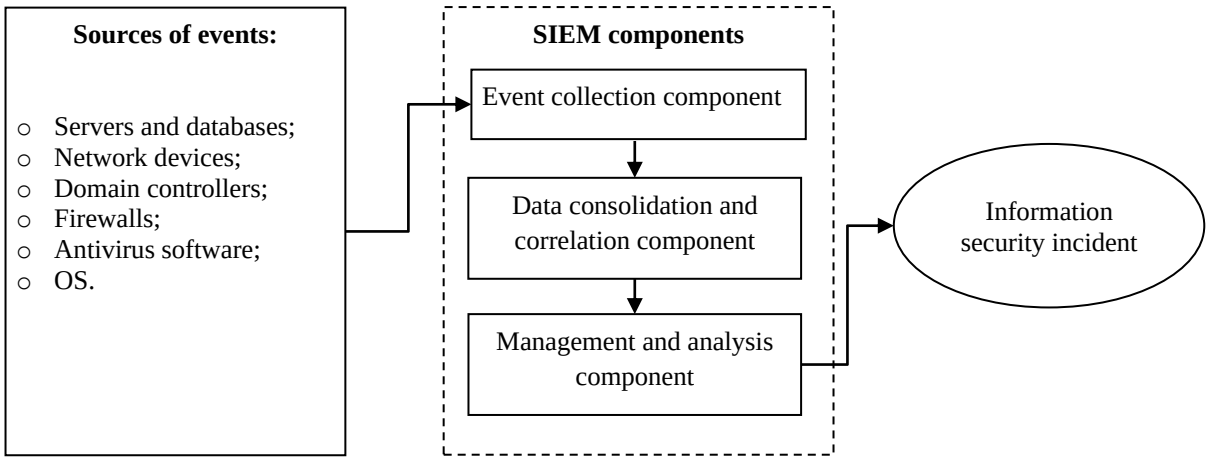

Fig. 2. Diagram of the SIEM system operation

Without the help of a SIEM, the amount of time it would take for security analysts to reliably detect suspicious activity by correlating logs between different types of devices would be very long given the complexity of most networks. It is rarely possible to detect and respond to any threat or attack on their infrastructures in time to prevent any damage. In addition, a SIEM solution can expand the possibilities of using the collected information [3].

It is clear that AI will become increasingly important in the future of SIEM as cognitive capabilities will improve the system's decision-making ability. It will also allow systems to adapt and evolve as the number of endpoints increases. As IoT, cloud computing, mobile, and other technologies increase the amount of data that a SIEM tool must consume. AI offers the potential for a solution that supports more types of data and a comprehensive understanding of threats as they evolve.

4. Anti-virus solutions based on AI

Artificial intelligence plays a key role in modern anti-virus solutions, such as Microsoft Defender Advanced Threat Protection (ATP) [11] and Darktrace [12]. These systems use AI capabilities to improve detection and response to cyber threats.

Microsoft Defender ATP [11] is an enterprise endpoint security platform designed to prevent, detect, investigate, and respond to threats. It integrates endpoint behavioral sensors, cloud-based security analytics, and threat intelligence to provide comprehensive protection. The sensors built into Windows 10 collect and process behavioral signals that are sent to the Microsoft Defender isolated cloud environment for further analysis.

The main features of Microsoft Defender include [11]:

- Reducing the attack surface: minimizing attack vectors to reduce the possibility of intrusion.
- Next-generation defense: using advanced machine learning techniques to detect sophisticated attacks.
- Endpoint Detection and Response (EDR): Provides in-depth analysis and visualization of threats in real time.
- Automated investigation and response: reduce the workload of security teams by automating processes.

Among the advantages of the Microsoft Defender solution is that it has deep integration with Windows, Microsoft 365, and Azure, but at the same time, this can be seen as a disadvantage, as it does not cover all network traffic as effectively as, for example, Daktrace [12], which will be discussed further. Other advantages include automatic threat research and built-in SIEM and SOAR capabilities through Microsoft Sentinel. Let's take a closer look at the Daktrace solution.

Darktrace [12] is a cybersecurity company that uses artificial intelligence and machine learning to detect cyberattacks and vulnerabilities in computer systems. Its technology aims to detect threats quickly and efficiently by using behavioral analysis to identify anomalies that may indicate attacks.

Darktrace uses self-learning AI to provide proactive cyber defense. Darktrace's core principle is the use of machine learning to analyze user, device, and network behavior. Its technology is capable of:

- Detect threats in real time: by analyzing network and user behavior to detect anomalies. And just as importantly, it detects new, unknown threats without using known algorithms or databases of known viruses.
- Autonomously respond to threats: perform automatic actions to neutralize attacks without human intervention.
- Provide protection across multiple environments: networks, email, cloud services, operational technologies and endpoints, as well as external platforms such as AWS, Azure, Google Cloud.

Thus, we can see that both Microsoft Defender for Endpoint [11] and Darktrace [12] use artificial intelligence (AI) to improve cybersecurity, but they have different approaches and areas of application. Let us consider them in a comparative analysis in Table 1.

Table 1

Comparative analysis of modern AI-based anti-virus solutions

Characteristics	Microsoft Defender ATP	Darktrace
Solution type	Endpoint Detection and Response (EDR), SIEM/SOAR	Network Detection and Response (NDR), Autonomous AI security
The main approach	Use behavioral analysis and threat intelligence to protect endpoints	Self-learning AI to analyze anomalies in the network, email, and cloud services
Main features	Endpoint threat detection, attack analysis, automated response	Network threat detection, autonomous response with Darktrace Antigena
Protection against ransomware	Blocks known attacks, analyzes behavior, stops suspicious processes	Detects anomalous activity in real time, blocks malicious activity at the network level
Focus on threats	Viruses, malware, exploits, attack scripts, account compromise	Anomalies in network traffic, insider threats, zero-day attacks
Integration	Deep integration with the Microsoft 365 ecosystem, Azure, Defender XDR	Support for hybrid environments: network, email, cloud, industrial systems
Cloud support	Microsoft 365, Azure	AWS, Google Cloud, Azure, local area networks

Let's consider who each solution is best suited for.

Microsoft Defender for Endpoint:

- Large and medium-sized companies operating in the Microsoft ecosystem.
- Organizations looking for powerful endpoint protection with built-in SIEM analytics.
- Companies with an IT team that is ready to manage security through Microsoft Security Center.

Darktrace would be better integrated into:

- Businesses with extensive networks that require in-depth traffic analysis.
- Organizations that want to detect anomalies in real time and automatically respond to them.
- Companies that use mixed environments (local servers, cloud platforms, IoT).

These examples demonstrate how the integration of AI into antivirus solutions improves the efficiency of detecting and responding to modern cyber threats, ensuring proactive protection of information systems.

5. AI as an attacker's tool – risks and threats associated with AI

The previous sections have discussed the benefits of using AI in cybersecurity, but there are undoubtedly many challenges to its use, as AI not only opens up new opportunities for development, but also becomes a tool in the hands of attackers, creating new risks and threats. The most common of these threats in the modern world are:

- AI-based attacks: Cybercriminals can use AI to launch more sophisticated and targeted attacks. This increases the effectiveness of their actions and makes it more difficult to detect threats.
- Disinformation and deep fakes (Deepfake): AI allows for the creation of realistic fake video, audio, and text that can undermine trust in authoritative sources, manipulate public opinion, and interfere with electoral processes.
- Automation of cyberattacks: Attackers can use AI to automate attacks, allowing them to be carried out faster and less likely to be detected.
- Attacks on AI systems: Attackers can manipulate the data that AI systems are trained on, leading to incorrect decisions and increasing the vulnerability of the systems.

Attackers are using AI to automate and increase the effectiveness of social engineering attacks. For example, neural networks can automatically generate highly believable phishing messages that convince users to share their passwords or other important information. This allows attackers to

gain access to the system without the need to conduct technical attacks, bypassing many levels of protection. Attackers are also actively using AI to carry out various attacks, which requires security professionals to develop new protection strategies. Therefore, let's take a look at the main types of AI-based attacks.

5.1. Data poisoning attacks

The first Data Poisoning attacks [15] were carried out in cybersecurity back in 2006 [19] and 2008 [20] and have since gained popularity among attackers. These attacks occur at the training or retraining stage of an AI model. Attackers inject malicious data into the training database, which leads to malfunctioning of the system and generation of false results. This can cause errors in classification or decision-making. For example, GANs can create artificial data that looks legitimate but is intended to mislead or corrupt machine learning models, which affects their performance and reliability.

Data poisoning can also reinforce existing biases in AI systems [16]. Attackers can target specific subsets of data, such as a particular demographic, to inject biased data. This can cause an AI model to perform unfairly or inaccurately. For example, facial recognition models trained with biased or fake data may incorrectly identify people from certain groups, leading to discriminatory results. These types of attacks can affect both the fairness and accuracy of ML models in different applications.

Data poisoning can also open the door to more sophisticated attacks [16], such as inversion attacks, in which hackers attempt to reconstruct the model's training data. Once an attacker successfully poisons the training data, they can then use these vulnerabilities to launch more serious attacks. In systems designed for sensitive tasks, such as cybersecurity, these risks can be particularly dangerous.

To protect against data poisoning attacks, organizations can implement strategies to help ensure the integrity of training datasets, improve model reliability, and continuously monitor AI models.

5.2. Evasion attacks

Evasion Attacks [18] consist of creating special input data that misleads the AI model, forcing it to make incorrect predictions or classifications. Such attacks can be aimed at bypassing malware detection systems or other protective mechanisms.

The discovery of evasion attacks against machine learning models has sparked increased interest in adversarial machine learning, leading to significant growth in this research area over the past decade. In an evasion attack, the goal of the attacker is to create competitive examples, which are defined as test samples [21].

In cybersecurity applications, competitive examples must respect the constraints imposed by the program semantics and the representation of cyber data features, such as network traffic or program binaries [18].

FENCE is a general framework for creating white-box evasion attacks using gradient optimization in discrete domains and supports a number of linear and statistical characteristic dependencies [22]. FENCE has been applied to two network security applications: malicious domain detection and malicious network traffic classification. In [23], this technique was applied to network intrusion detection and phishing classifiers. It is noted in [23] that continuous domain attacks cannot be easily applied in constrained environments because they result in infeasible adversarial examples. Pierazzi et al. [24] discuss the difficulty of establishing possible evasion attacks in cybersecurity due to constraints in the function space and formalize evasion attacks in the problem space and create possible adversarial examples for Android malware.

5.3. Rapid deployment attacks

In a Prompt Injection attack [18], attackers insert malicious instructions into requests to AI models, forcing them to perform unwanted actions or provide sensitive information, i.e., tricking the model into returning an unexpected response and causing the application to act in unplanned ways.

Successful implementation can lead to the leakage of confidential data, destruction of information, and other types of damage depending on the application.

5.4. Social engineering attacks using AI

AI is used to automate and improve the effectiveness of social engineering attacks, such as phishing or manipulation, making them more difficult to detect.

Attackers use social engineering techniques to conceal their true "identity" by posing as trusted organizations or individuals to victims. These attacks are aimed at obtaining personal information to access the target network through deception and manipulation. Social engineering is used as the first stage of a large cyberattack to penetrate a system, install malware, disclose confidential data, etc.

For example, in the case of phishing [18], it was previously demonstrated that large language models (LLMs) can create persuasive scams such as phishing emails [25]. Now that LLMs can more easily integrate with applications, they can not only create fraudulent activities, but also widely distribute such attacks [26]. Users are likely to be more susceptible to these new attacks, as opposed to phishing emails, because they lack experience and awareness of this new threat technique.

The LLM itself also acts as a computer on which malicious code runs and spreads. For example, an automated message processing tool that can read and create emails and view users' personal data can propagate the injection to other models that can read these incoming messages [26].

So, let's look at measures to counter AI-based attacks. To protect yourself from AI attacks, you need to take the following steps:

- Improving data quality: Ensuring data is clean and reliable for training AI models helps reduce the risk of data poisoning.
- Developing resilient models: Creating AI models that are resistant to attacks by implementing security methods and regular testing.
- Monitoring and auditing: Continuously monitor the operation of AI models and conduct audits to identify possible vulnerabilities.
- Staff training: Raising employees' awareness of possible AI-based attacks and methods of detecting them.

In today's digital environment, it is important to be prepared for new challenges related to the use of AI and implement appropriate cybersecurity measures.

Conclusions

1. This paper reviews and analyzes. This article provides a comprehensive analysis of the current state and prospects of artificial intelligence (AI) in cybersecurity. Both the benefits of implementing AI in security systems and the risks associated with its use are considered.

2. AI allows you to automate the process of detecting and responding to threats, which significantly increases the effectiveness of cyber defense. The use of machine learning algorithms helps to quickly analyze large amounts of data and identify anomalies in the behavior of users and systems.

3. Modern cybersecurity systems, such as SIEM (Security Information and Event Management), benefit greatly from AI integration, as they are able to analyze events in real time and warn of possible attacks. Continuously learning from the past improves the accuracy and reliability of AI-powered SIEM systems against increasingly powerful cyber threats. Ultimately, an AI-powered SIEM integrates various components such as AI, ML, deep learning, NLP, and UEBA. This integration leads to smarter, more efficient and proactive cybersecurity measures, which is crucial in the ever-changing cyber threat environment. At the same time, a large number of integrations with various systems allow SIEM systems to monitor and accumulate data on the current state of cybersecurity of information infrastructure in relation to certain international and national standards, such as ISO 27001, GDPR or PCI DSS.

4. AI can analyze huge amounts of data and identify patterns that indicate potential threats, allowing organizations to stay ahead of hackers. AI can also identify threats faster and more accurate-

ly, and automatically block malicious traffic without human intervention. Thanks to behavioral analysis, AI antiviruses (e.g., Microsoft Defender ATP and Darktrace) can detect threats more effectively than traditional antivirus programs.

5. As for the use of attack detection protection systems, if a company actively uses Microsoft 365, Azure, and Windows, it is better to choose Microsoft Defender for Endpoint. It provides deep integration, automated response, and behavioral analysis of threats on endpoints. If you need flexible and autonomous protection of all levels of your network, you should consider Darktrace. It is suitable for organizations that want to detect anomalies, analyze cyber threats in real time, and respond to them without human intervention. But the ideal option is to combine both solutions: Microsoft Defender for Endpoint for endpoint protection and Darktrace for network analysis and automated threat response.

6. The use of AI not only for defense but also for attack poses a significant threat. Attackers can use AI to automate attacks, manipulate, and engage in social engineering. Attackers are increasingly using AI to create sophisticated malware that can adapt to defenses. Security experts are worried about potential autonomous AI attacks, forcing companies to prepare now. Organizations need to implement comprehensive strategies to take advantage of the benefits of AI while minimizing its potential threats.

7. The prospects for development and recommendations for the use of AI in cybersecurity are as follows:

- Increasing the transparency of AI algorithms and implementing ethical standards are critical to the credibility of AI technologies in the security sector.
- Development of stable AI models that will be less vulnerable to attacks by malicious actors.
- Investing in research on post-quantum cryptography and the latest authentication methods to strengthen cybersecurity systems.
- Strengthening international cooperation in the field of AI standards and regulation, in particular through the initiatives of organizations such as NIST.

8. Thus, artificial intelligence has the potential to completely change the approach to cybersecurity. Its ability to quickly analyze large amounts of data, predict threats, and automatically respond to attacks makes it a key element of protection in the digital world. However, it should be remembered that cybercriminals are also using AI, so the future of cybersecurity will depend on the balance between security innovations and threats from criminal groups.

9. The future of cybersecurity is a symbiosis of humans and artificial intelligence, where analysts and security experts collaborate with smart systems to create a secure cyberspace.

References:

1. NIST standardization process "Post-Quantum Cryptography: Digital Signature Schemes". Access mode: https://csrc.nist.gov/Projects/pqc-dig-sig/round-1-additional-signatures.

2. TAO, Feng; Akhtar, Muhammad Shoaib; Jiayuan, Zhang. The future of artificial intelligence in cybersecurity: A comprehensive survey // EAI Endorsed Transactions on Creative Technologies. 2021. 8.28: e3–e3. https://doi.org/10.4108/eai.7-7-2021.170285.

3. Leung B. K. (2021). Security Information and Event Management (SIEM) Evaluation Report. ScholarWorks. May 2021. Access mode: https://scholarworks.calstate.edu/downloads/41687p49q.

4. González-Granadillo G., González-Zarzosa S., Diaz, R. Security Information and Event Management (SIEM) // Analysis, Trends, and Usage in Critical Infrastructures. Sensors. 2021. 21(14). Access mode: https://doi.org/10.3390/s21144759

5. Muhammad S., et al. Effective Security Monitoring Using Efficient SIEM Architecture // Human-centric Computing and Information Sciences. 2023. 13. Access mode: https://doi.org/10.22967/HCIS.2023.13.017.

6. What is SIEM. Security Information and Event Management Tools. (n.d.). Imperva. Access mode: https://www.imperva.com/learn/application-security/siem/.

7. IBM Security QRadar. What is security information and event management (SIEM)? https://www.ibm.com/think/topics/siem.

8. Splunk. The Splunk SIEM. Access mode: https://www.splunk.com/en_us/products/enterprise-security.html.

9. Stellar Cyber. AI SIEM: The 6 Components of AI-Based SIEM. - Access mode: https://stellarcyber.ai/learn/ai-driven-siem/.

10. ISO/IEC 27001:2022. Information technology - Security techniques - Information security management systems - Requirements. International standard. 3 Edition.
11. Microsoft Defender for Endpoint. 2024. Access mode: https://learn.microsoft.com/uk-ua/defender-endpoint/microsoft-defender-endpoint.
12. Darktrace. Official website. 2025. Access mode: https://darktrace.com/.
13. Mauri L., Damiani E. Modeling Threats to AI-ML Systems Using STRIDE. *Sensors* 2022, *22*(17), 6662; - Access mode: https://doi.org/10.3390/s22176662.
14. The near-term impact of AI on the cyber threat: https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat.
15. Nihad Hassan. What is data poisoning (AI poisoning) and how does it work? Search Enterprise AI, TechTarget, 2024. Access mode: https://www.techtarget.com/searchenterpriseai/definition/data-poisoning-AI-poisoning.
16. Tom Krantz, Alexandra Jonker. What is data poisoning? IBM. Access mode: https://www.ibm.com/think/topics/data-poisoning.
17. NIST Trustworthy and Responsible AI NIST AI 100-5. A Plan for Global Engagement on AI Standards: https://doi.org/10.6028/NIST.AI.100-5.
18. Vassilev A, Oprea A, Fordyce A, Anderson H (2024) Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations. (National Institute of Standards and Technology, Gaithersburg, MD) NIST Artifcial Intelligence (AI) Report, NIST Trustworthy and Responsible AI NIST AI 100-2e2023. Access mode: https://doi.org/10.6028/NIST.AI.100-2e2023.
19. R. Perdisci, D. Dagon, Wenke Lee, P. Fogla, and M. Sharif. Misleading worm signature generators using deliberate noise injection // 2006 IEEE Symposium on Security and Privacy (S&P'06), Berkeley/Oakland, CA, 2006.
20. Blaine Nelson, Marco Barreno, Fuching Jack Chi, Anthony D. Joseph, Benjamin I.P. Rubinstein, Udam Saini, Charles Sutton, and Kai Xia. Exploiting machine learning to subvert your spam flter // First USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET 08), San Francisco, CA, April 2008. USENIX Association.
21. Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks // International Conference on Learning Representations, 2014.
22. Alesia Chernikova and Alina Oprea. FENCE: Feasible evasion attacks on neural networks in constrained environments // ACM Transactions on Privacy and Security (TOPS) Journal. 2022.
23. Ryan Sheatsley, Blaine Hoak, Eric Pauley, Yohan Beugin, Michael J. Weisman, and Patrick McDaniel. On the robustness of domain constraints // Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, CCS '21, p. 495–515, New York, NY, USA, 2021. Association for Computing Machinery.
24. Fabio Pierazzi, Feargus Pendlebury, Jacopo Cortellazzi, and Lorenzo Cavallaro. Intriguing properties of adversarial ML attacks in the problem space // 2020 IEEE Symposium on Security and Privacy (S&P). P. 1308–1325. IEEE Computer Society, 2020.
25. Daniel Kang, Xuechen Li, Ion Stoica, Carlos Guestrin, Matei Zaharia, and Tatsunori Hashimoto. Exploiting programmatic behavior of llms // Dual-use through standard security attacks. arXiv preprint arXiv:2302.05733, 2023.
26. Kai Greshake, Sahar Abdelnabi, Shailesh Mishra, Christoph Endres, Thorsten Holz, and Mario Fritz. Not what you signed up for // Compromising realworld llm-integrated applications with indirect prompt injection. arXiv preprint arXiv:2302.12173, 2023.

Received 08.02.2025

Information about the authors:
Yuriy Golikov – CEO and Founder of DevBrother tech company, USA; e-mail: yuriy@devbrother.com ; ORCID: https://orcid.org/0009-0008-7946-4663