ASEJ ISSN: 2543-9103 ISSN: 2543-411X (online)

AI in cybersecurity

Konrad Hoza[1]

[1]Silesian University of Technology
Poland

***Abstract*— This comprehensive review of the use of artificial intelligence in cybersecurity highlights its significant potential in threat detection, response, and overall security posture. Although significant progress has been made in using artificial intelligence in cybersecurity, future research must focus on optimizing specific AI techniques, improving interpretation, and solving problems related to implementation in different environments. Further development of AI methodologies and their applications can contribute to increasing resilience and adaptability in the field of cybersecurity in the face of evolving threats.**

Keywords— artificial intelligence, AI, cybersecurity, machine learning

I. INTRODUCTION

We are more present in the online world than ever before. Access to the Internet for the average person living in Europe is almost as high as access to electricity or clean water. Such a high level of access to the network in the present times has made it the main tool for communication, work, learning, and access to information. However, the Internet, despite its many advantages, is not without its drawbacks. Virtual space, despite its usefulness, is also a place where dangers lurk for people who do not have sufficient knowledge about safe browsing. Children, the elderly, and people less familiar with technology are particularly at risk, but anyone can be affected. Therefore, cybersecurity means not only secure websites and connections, but also safe and conscious use of access to freedom on the web.

II. ARTIFICIAL INTELLIGENCE

In its broadest defnition, AI is equated with algorithms. However, this is not an especially useful approach for our analysis. Algorithms predate AI and have been widely used outside this feld. The term ‘algorithm’ is derived from the name of the ninth-century Persian mathematician Mohammed ibn Musa al-Kharizmi and refers to a specifc instruction for solving a problem or performing a calculation. If we were to defne AI simply as the use of algorithms, it would include many other activities such as the operations of a pocket calculator or even the instructions in a cookbook. In its strictest defnition, AI stands for the imitation by computers of the intelligence inherent in humans. Purists, however, point out that many current applications are still relatively simple and therefore not true AI (Sheikh, et al., 2023).

Artificial intelligence is one of the most groundbreaking technologies of our time, but it is worth noting that a legal definition of artificial intelligence has not yet been developed, both in national legislation and in international conventions. In European Commission documents, artificial intelligence is defined as: "a machine-based system that is able to influence the environment by generating output data (predictions, recommendations or decisions) for a given set of goals.

It uses input data from machines or people to:

perceive real or virtual environments;

translate these observations into models through automated analysis (e.g. using machine learning) or manually;

use model inference to formulate outcome options.

Artificial intelligence systems are designed to operate with varying levels of autonomy" (European Commission, 2019).

III. TECHNIQUES AND METHODS OF MACHINE LEANING IN ARTIFICIAL INTELLIGENCE

Machine learning is an integral part of the Artificial Intelligence revolution. Unlike traditional programming based

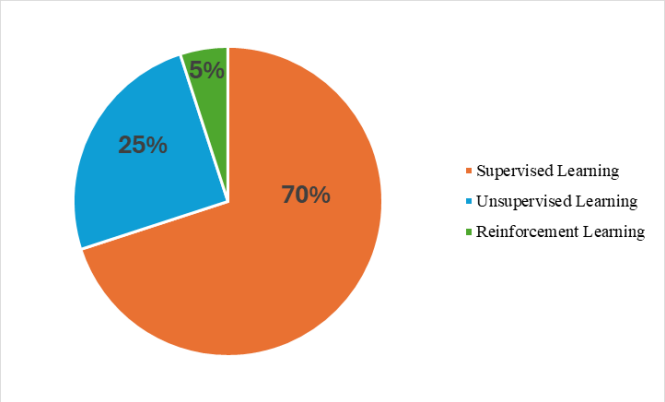
ASEJ - Scientific Journal of Bielsko-Biala School of Finance and Law
Volume 29, No 3 (2025), pages 5
https://doi.org/10.19192/wsfip.sj3.2025.4
Received: March 20025, Accepted: September 2025,
Published: September 2025

on rigid, human-defined rules, machine learning uses a probabilistic approach. In this setup, the models learn patterns and relationships directly from the data. Artificial intelligence uses the knowledge and conclusions that machine learning has drawn from processed data sets after analyzing them using machine learning algorithms. These systems are characterized by the ability to continuously optimize and adapt in response to incoming streams of new data. This, in turn, allows for maintaining a high level of consistency and accuracy in models in dynamically changing environments without the need for costly reorganization and reimplementation of the system. Machine learning is therefore an indispensable part of artificial intelligence and its development has a corresponding impact on AI. Consequently, further development of artificial intelligence is inextricably linked to progress in the field of machine learning. Each machine learning model and algorithm fits into one of these categories based on its approach to learning and problem-solving.

Machine learning algorithms are divided into three categories (Figure 1):

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

FIGURE 1. QUALIFICATION OF MACHINE LEARNING MODELS



Source: (Loza, 2025)

Supervised learning - controlled machine learning models learn from labeled data, which includes both input data and the corresponding output. Each data point in the training set contains input features (e.g., email subject, email content, sender address) and a target label (e.g., "spam" or "not spam"). The model learns to map input data to output data by minimizing the difference between predicted and actual labels thereby enabling it to classify new, unseen data. In the field of cybersecurity, supervised learning is widely used for filtering spam, classifying malware, and detecting phishing, where a large amount of labeled training data is available.

Unsupervised learning - unsupervised machine learning models learn from data sets containing only input features without defined output labels. Such data sets may include network traffic logs, system event records, or metadata of file attributes. The goal is not to predict a specific label, but to discover hidden patterns, groupings, or anomalies in the data. Typical tasks include grouping, dimensionality reduction, and anomaly detection. In the field of cybersecurity, unsupervised learning is particularly useful in intrusion detection systems (IDS), which work on the principle of anomaly detection — they identify deviations from normal network behavior as potential indicators of malicious activity.

Reinforcement learning - involves training an agent to make sequential decisions by interacting with the environment and receiving feedback in the form of rewards or penalties. Unlike supervised learning, the optimal action in a given state is not clearly defined; the agent must explore and adapt strategies to maximize total rewards over time. The input to the agent represents the current state of the system (e.g., network configuration, detected traffic patterns), while the output responds to actions (e.g., blocking suspicious IP addresses, isolating infected hosts). In the field of cybersecurity, Reinforcement Learning can be used to develop adaptive defense systems, such as automatically adjusting the firewall, dynamically deploying traps, or resource allocation strategies in the event of incident response.

IV. AI AND CYBERSECURITY – PROBLEMS

One of the key areas of AI application is cybersecurity. Artificial intelligence-based systems can analyze huge amounts of network data in real time, detect anomalies and potential threats, and automatically respond to incidents. This allows security teams to deal more quickly and effectively with the increasingly complex landscape of cyber threats.

Machine learning and deep learning algorithms are used to analyze network traffic in real time, which allows for the detection of unusual behavior and the blocking of previously unknown threats (Okdem, 2024).Their ability to analyze data can also be used after an attack to find correlations between the attacked files and thereby contribute to increasing future protection. Furthermore, natural language processing techniques support automatic analysis of malware code, which enables classification of new malware variants. It is also worth mentioning that artificial intelligence is used to aggregate data from various sources of threats, for example honeypots that lure cybercriminals to examine their methods of attack, gather information, and improve system security based on that information. A hacker attack on such a system has no chance of success, as it is carefully isolated from the production network so that in the event of a successful attack, the data is not damaged. When an attacker enters a honeypot, all of their actions are monitored, recorded, and analyzed, which enables the construction of predictive threat models and dynamic security adaptation (Akhtar, Feng, 2022).

In recent years, there has been a dynamic increase in the implementation of artificial intelligence tools in the area of cybersecurity, both in the private and public sectors. Solutions like Darktrace use unsupervised machine learning to model normal network behavior and detect anomalies in real-time, enabling proactive identification of zero-day threats. Cylance, on the other hand, uses supervised learning models based on the analysis of binary file features, allowing it to detect malware before it is even launched, which significantly reduces the time

it takes to respond to incidents. The Microsoft Defender AI solution integrates natural language processing (NLP) and behavioral analysis techniques, which enable fast correlation of events from multiple sources and automatic generation of recommendations for corrective actions. In the financial sector, these tools are used for transaction fraud detection and electronic banking system protection, while in public administration, they are used for monitoring critical infrastructure and responding to ransomware attacks. Research results indicate that integrating AI into security processes can increase incident detection effectiveness by 30–50% compared to traditional methods (Ofusori et al., 2024; Welukar et al., 2021).

One of the main problems with implementing AI in cybersecurity is generating false alarms. The system may consider harmless activity as a potential attack due to the model's excessive sensitivity or incorrect patterns in the training data.

Example:

At the financial institution, the SI intrusion detection system triggered alarms whenever a remote login attempt was made from outside the country. Although some of the attempts were indeed suspicious, most of them came from employees working remotely during business trips. As a result, the security team lost several hours a day verifying false incidents. As with the constant false alarms, it led to a decrease in the alertness of the staff, which increased the risk of missing a real attack.

Large and representative data sets are needed. Machine learning models in cybersecurity must be trained on large, diverse, and current data sets to effectively identify both known and new types of threats. Lack of such data results in the algorithm's limited ability to recognize unknown attack patterns.

Example:

During the development of an AI system for detecting ransomware attacks, the research team encountered the problem of limited samples of actual attacks. It was necessary to generate artificial data simulating the behavior of ransomware, which, however, reduced the effectiveness of the model in real conditions — the algorithm recognized only attacks with a similar pattern to those it was trained on.

The „black-box" problem: Many AI algorithms, especially those based on deep neural networks, operate in a way that is difficult to interpret. Lack of transparency in the decision-making process raises doubts about the credibility of the detected incidents and makes it difficult to meet legal requirements. Systems like Darktrace or Cybereason can autonomously detect anomalies, but often do not provide detailed justifications for their decisions (Barredo et al., 2020).

For example: In theory, an application flagged as "unusual" could be automatically blocked from functioning without the system explaining which specific behavior (e.g., access to the registry, an unusual network connection) triggered the alarm. This problem would be a challenge for employees, as models based on deep neutron networks do not explain why they made a given decision. Therefore, instead of shortening the reaction time, it may even be prolonged in some cases.

V. BENEFITS OF AI IN CYBERSECURITY

An examination of the advantages of artificial intelligence in the field of cybersecurity reveals that institutions that have implemented AI reap significant benefits. This is evident because the return on investment for a significant number of organizations has increased when they adopted cybersecurity tools.

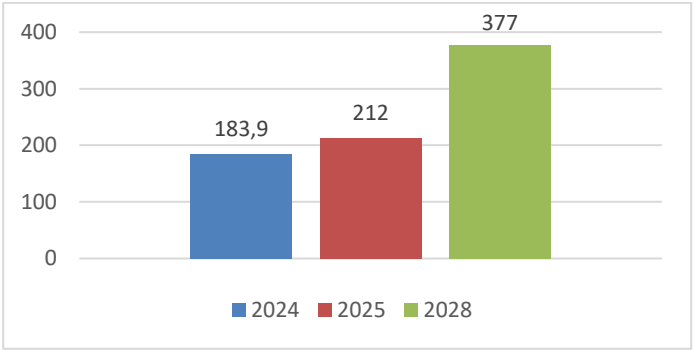
Example:

Siemens AG, a global leader in electrification, automation, and digitalization, struggled with the problem of overwhelming alerts from all over the world on a daily basis. Employees had to monitor the environment worldwide as new, previously unheard-of threats emerged. In response to these threats, Siemens AG used Amazon Web Services (AWS) to create an artificial intelligence-based, fast, self-monitoring, and extremely flexible platform for its Siemens Cyber Defense Center (CDC).

The use of artificial intelligence in cybersecurity allows institutions to understand and reapply previous threat patterns to identify new threats. This saves time and effort in identifying and investigating incidents and fixing threats. Artificial intelligence offers opportunities in the field of cyber security, mainly because the cyber security landscape is rapidly moving from identification, manual response, and mitigation to automatic mitigation. Artificial intelligence can identify new and complex modifications in the scope of attack extensibility.

The future: As the market becomes increasingly aware of the cyber threats that lurk, spending on network security will increase in the coming years. Gartner predicts a significant 15% increase in global cybersecurity spending in 2025, reaching $212 billion, compared to $183.9 billion in 2024 (Gartner, Stamford, 2024). This growth is primarily driven by the security services segment, followed by security software and network security. A particularly dynamic growth is expected in the area of security software, which is strongly linked to the adoption of artificial intelligence (AI) and generative AI (GenAI) in the security of applications, data, and infrastructure. IDC has looked even further into the future and predicts that by 2028, spending on network security will exceed double the value of 2024, reaching $377 billion (Tennyso, Perini, 2025).

FIGURE 2. WORLDWIDE ESTIMATED CYBERSECURITY SPENDING (IN BILLIONS USD), 2024–28



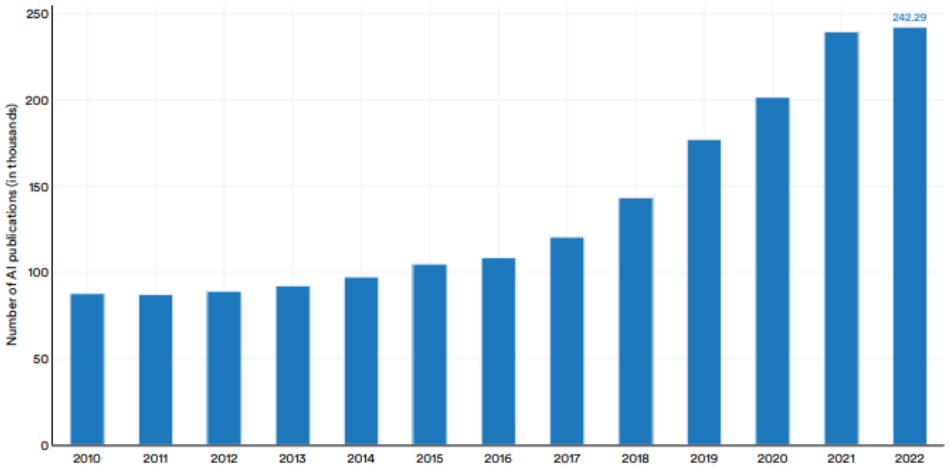
Source: own study based on Garthner & IDC works.

The numbers show a significant and steady increase in the total number of publications on artificial intelligence, which highlights the rapid expansion of this field and its growing

importance in the global scientific environment. This growth is potentially the result of increased investment, technological advancements, and recognition of AI's transformative potential in various sectors.

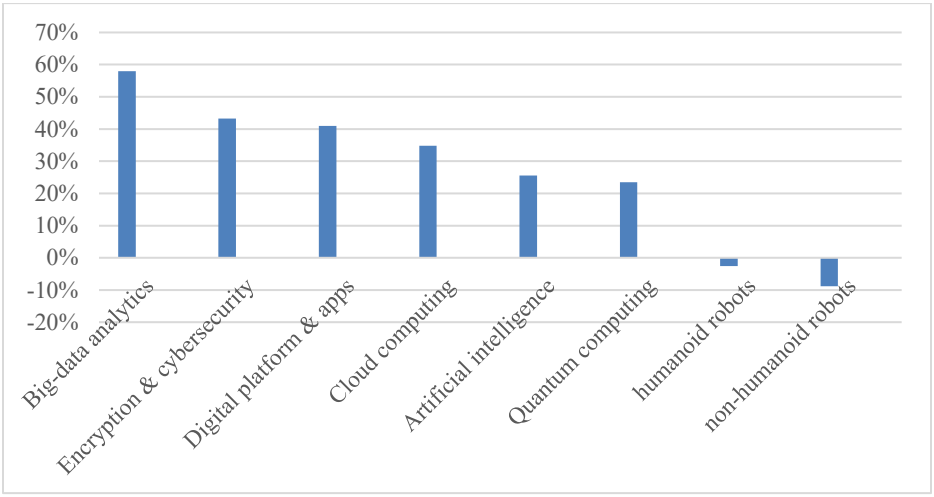
Also, data on publication channels show significant evolution. The growing share of peer-reviewed articles in journals, which may exceed the number of papers presented at conferences, indicates a field that is consolidating its knowledge base. This may reflect a trend towards more detailed, verified, and long-term research (Figure 3).

FIGURE 3. NUMBER OF AI PUBLICATIONS IN THE WORLD, 2010–22



Source: Stanford University, 2024

FIGURE 4. EXPECTED IMPACT OF TECHNOLOGY TO CREATE OR DISPLACE JOBS, 2023–2027



Source: own study based on Future of Jobs Report 2023

SI offers us automation of processes, improvement of work efficiency, fast data processing, as well as development of new groundbreaking services, such as medical diagnostics support systems or intelligent assistants. Artificial intelligence poses a threat to certain professions, such as accountants, planning assistants, cashiers, truck drivers, and many others. Such changes can therefore exacerbate social inequalities, raise ethical concerns related to responsibility or privacy (Global Workforce, 2025).

According to the research, the study also analyzes the expected impact of adopting technology on employment (Figure 4) shows that all technologies, except two, will create new jobs over the next five years. It is expected that analyzing large data sets, climate and environmental change management technologies, and encryption and cybersecurity will be the biggest drivers of employment growth (Future of Jobs Report 2023). Integrating traditional tools with AI models significantly increases the level of security, reduces costs, and saves both time and money. A significant advantage of AI is the automation of reactions, which allows it to take action on its own to stop an attack, prevent it, or minimize the damage caused – for example, by isolating already infected systems from the rest. This significantly speeds up response time and reduces the burden on employees.

VI. SUMMARY

Understanding artificial intelligence not only as an algorithm processing data or commands, but as a set of techniques and methods of knowledge modeling, included in a system and subject to interactions with the external environment, it should be stated that the rules and procedures of cybersecurity are no longer sufficient for building and maintaining the resilience of

the AI system itself, as well as the resilience of the environment in which it is embedded.

This comprehensive review of the use of artificial intelligence in cybersecurity highlights its significant potential in threat detection, response, and overall security posture. Although significant progress has been made in using artificial intelligence in cybersecurity, future research must focus on optimizing specific AI techniques, improving interpretation, and solving problems related to implementation in different environments. Further development of AI methodologies and their applications can contribute to increasing resilience and adaptability in the field of cybersecurity in the face of evolving threats.

VII. REFERENCES

Akhtar, Muhammad & Feng, Tao. (2022). Malware Analysis and Detection Using Machine Learning Algorithms. Symmetry. 14. 2304. 10.3390/sym14112304

Barredo A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

European Commission: Directorate-General for Communications Networks, Content and Technology and Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Ethics guidelines for trustworthy AI, Publications Office, 2019, https://data.europa.eu/doi/10.2759/346720

Gartner, Stamford, (2024) https://www.gartner.com/en/newsroom/press-releases/2024-08-28-gartner-forecasts-global-information-security-spending-to-grow-15-percent-in-2025

How Will AI Affect the Global Workforce? (2025) https://www.goldmansachs.com/insights/articles/how-will-ai-affect-the-global-workforce

Loza, B (2025) Supervised Machine Learning in Cybersecurity: A Comprehensive Analysis https://medium.com/@leev574/supervised-machine-learning-in-cybersecurity-a-comprehensive-analysis-04ac97a822fc

Okdem, S. (2024) Artificial Intelligence in Cybersecurity: A Review and a Case Study. Appl. Sci. 14, 1 https://doi.org/ 10.3390/app142210487

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

Sheikh, H., Prins, C., Schrijvers, E. (2023). Artificial Intelligence: Definition and Background. In: Mission AI. Research for Policy. Springer, Cham. https://doi.org/10.1007/978-3-031-21448-6_2

Stanford University, (2024) The 2024 AI Inder Report, https://hai.stanford.edu/assets/files/hai_ai-index-report-2024-smaller2

Tennyson M., Perini S., (2025), IDC, https://my.idc.com/getdoc.jsp?containerId=prEUR253264525

Welukar, J., Gagan, P. (2021) Artificial Intelligence in Cyber Security - A Review, International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 6, pp. 488-491, November-December 2021. Available at doi : https://doi.org/10.32628/IJSRST218675 Journal URL : https://ijsrst.com/IJSRST218675

World Economic Forum, (2023) Future of Jobs Report 2023 https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023