Intelligence Defense - the Role of Artificial Intelligence in Combating Cyber Threats

Silvia OVREIU, Daniel SAVU, Electra MITAN, Delia NEACȘU
National Institute for Research &Development in Informatics - ICI Bucharest
silvia.ovreiu@ici.ro, daniel.savu@ici.ro, electra.mitan@ici.ro, delia.radulescu@ici.ro

Abstract: Digital transformation has led to increased dependence on critical infrastructures and their exposure to cyber threats that traditional defense methods cannot effectively manage. This article highlights the use of artificial intelligence and machine learning in detecting, preventing, and responding to attacks, including zero-day attacks. Metrics for evaluating anomaly detection are presented, underlying that modern AI-driven anomaly detection systems must deliver high accuracy, precision, and recall, while minimizing false positives and detection latency. Also, AI-based security solutions and various AI technologies applied in cybersecurity are presented. In addition, the ethical, technical, and operational challenges of implementing AI in intelligent, explainable, and proactive cyber defense applications are highlighted.
Keywords: Cybersecurity, artificial intelligence, critical infrastructures, intelligent defense.

INTRODUCTION

Digital transformation has increased dependence on critical infrastructures (such as power grids, telecommunications, transport, and financial infrastructures). The increase in interconnectivity and automation has led to infrastructures exposed to cyber threats that exceed the capabilities of traditional defense methods. Classic security solutions can no longer cope with the attacks. Therefore, advanced technologies and strategies became necessary to adopt.

Globally, cybersecurity is grappling with increasingly complex challenges. Over the past 20 years, cyber-attacks have surged at an alarming rate, both in variety and volume. Advanced attacks, such as Advanced Persistent Threats (APT) (Kumari & Lee 2023) or the exploitation of IoT device networks (Franco et al., 2021), demonstrated that these are no longer isolated incidents but a global-level concern.

Traditional defense techniques, reliant on fixed rules and convolutional detection systems, have proven to be inadequate. These methods perform well against known cyber–attacks but struggle to identify emerging threats that evolve rapidly. These limitations become even more pronounced in environments where configurations, traffic, and resources shift quickly.

https://doi.org/10.54851/v8i1y202608

INTRODUCTION

Digital transformation has increased dependence on critical infrastructures (such as power grids, telecommunications, transport, and financial infrastructures). The increase in interconnectivity and automation has led to infrastructures exposed to cyber threats that exceed the capabilities of traditional defense methods. Classic security solutions can no longer cope with the attacks. Therefore, advanced technologies and strategies became necessary to adopt.

Globally, cybersecurity is grappling with increasingly complex challenges. Over the past 20 years, cyber-attacks have surged at an alarming rate, both in variety and volume. Advanced attacks, such as Advanced Persistent Threats (APT) (Kumari & Lee 2023) or the exploitation of IoT device networks (Franco et al., 2021), demonstrated that these are no longer isolated incidents but a global-level concern.

Traditional defense techniques, reliant on fixed rules and convolutional detection systems, have proven to be inadequate. These methods perform well against known cyber–attacks but struggle to identify emerging threats that evolve rapidly. These limitations become even more pronounced in environments where configurations, traffic, and resources shift quickly.

Today, Artificial Intelligence (AI) is pivotal in cybersecurity for detection, prevention, and response. Machine Learning (ML) and Deep Learning (DL) algorithms can discern subtle patterns and predict zero-day attacks. Furthermore, AI enables automated responses, significantly reducing the time required to block an attack. As a result, the integration of AI algorithms into cybersecurity is essential today.

The purpose of the article is to explore the role of artificial intelligence in enhancing cybersecurity by analyzing the main AI models used for detecting and preventing cyber-attacks. It aims to highlight the benefits AI offers over traditional methods, discuss the challenges and limitations of AI deployment in cybersecurity, and outline future research directions. The article also presents performance metrics for AI-based anomaly detection and showcases real-world AI-driven security solutions.

The remainder of this paper is structured as follows. The section „Theoretical Foundations: Key Concepts" provides the theoretical framework necessary to comprehend the subject matter. Following this, the section „Metrics for Evaluating Anomaly Detection in AI-Based Cyber Defense," discusses the metrics used to assess AI-based anomaly detection systems within the cyber defense landscape. Next, the focus to „AI-Based Security Solutions" where we examine AI techniques utilized in cybersecurity, showcase practical applications, and compare them with traditional methods. Section „AI Technologies in Cybersecurity," then investigates the specific AI technologies applied in this domain. The paper proceeds with an in-depth look at „Challenges and Limitations". This section delves into the constraints of AI technologies, the difficulties associated with their implementation in security, organizational, and operational issues that arise, and areas for future research. In „Examples of AI-based Platforms in Real-World Applications" are presented concrete examples of how AI-driven solutions enable real-time threat detection, fraud prevention, and efficient cybersecurity responses across organizations. The section dedicated to the future directions marks the shift from reactive to proactive, adaptive, and autonomous defense strategies, with a focus on explainability, resilience, and ethical considerations, for the development of trustworthy AI systems. Finally, the paper concludes with „Conclusions", which offers a comprehensive summary of the research.

THEORETICAL FOUNDATIONS: KEY CONCEPTS

In the digital age, cybersecurity is not merely an option but a necessity. It encompasses the protection of computer systems, networks, and data from unauthorized access, attacks, or information theft. The aim is to ensure that data remains confidential, intact, and accessible whenever required. This necessity is even more apparent in the context of cyber threats, which

include attacks or exploitation of software vulnerabilities. Cyber threats encompass all forms of attacks targeting computer systems, networks, data, or users. They are no longer infrequent occurrences but global phenomena impacting individuals, organizations, and even national security. Table 1 outlines various types of cyber threats.

Table 1: *Types of cyber threats*

Type of attack	Description
Phishing	It is a form of cyber fraud where crucial information, such as financial or personal data, is extracted via an email that appears trustworthy.
Malware	It refers to malicious software designed to corrupt a computer system's data or disrupt its functionality.
DoS/DDoS	This type of cyber-attack overwhelms a server by sending a vast number of requests, causing it to freeze or crash. The objective is to disrupt or halt the functionality of the targeted serve.
MitM	The Man-in-the-Middle (MitM) attack is a cyber-attack that intercepts or disrupts communication between two parties to steal private information.
SQL Injection	A cyberattack targeting a web application aimed at manipulating the database to gain unauthorized access to the system.
Ransomware	A type of malware attack that locks a user's system by encrypting it and demanding a ransom to restore access.
APT	Advanced Persistent Threat (APT) attacks are intricate assaults executed by a well-organized group that gains unauthorized access to a network and remains undetected for an extended duration.
Zero-day	cIt refers to a cyber-attack that exploits an unknown vulnerability in a system, software, or hardware.
Password attack	It is a cyber-attack aimed at guessing, cracking, or stealing the user's password, attempting to gain unauthorized access to a system.

Artificial Intelligence, Machine Learning, Deep Learning, and Cyber Threat Intelligence

AI is a branch of computer science focused on creating systems capable of mimicking human cognitive functions such as learning, reasoning, and decision-making. In cybersecurity, AI enables the automatic analysis of behaviors across large volumes of data (Adi, Baig & Zeadaly, 2022).

ML, a component of AI, enhances algorithms that autonomously learn from historical data without needing specific programming for each task. The most commonly used ML techniques include supervised learning algorithms like K-Nearest Neighbors (K-NN), Random Forest (RF), and Support Vector Machines (SVM), as well as unsupervised learning algorithms such as clustering (Sutton& Barto, 1998). These methods effectively detect anomalies and identify unknown attacks, including zero-day threats.

DL is an advanced subset of ML based on artificial neural networks that can process large datasets and identify details often overlooked by humans. Convolutional Neural Networks (CNN) are a specific type of artificial neural network used for image recognition and processing. They recognize complex patterns by automatically extracting features from data (LeCun, Bengio & Hinton, 2015).

Moreover, Generative AI (GenAI) is a subdomain of DL with the ability to produce new content, such as texts and images, aiding in the development of proactive cybersecurity strategies (Ferrag et al., 2025). In cybersecurity, DL and GenAI techniques are utilized to analyze user behavior and detect hard-to-identify attacks, such as APTs (see Table 1).

A key term in cybersecurity is Cyber Threat Intelligence (CTI), which involves the analysis, collection, and application of cyber threat data to avert attacks such as malware, APTs, or zero-day attacks. When integrated with AI, CTI facilitates rapid, preventive, and adaptive responses to attacks (Ainslie et al., 2023). Together, these technologies form an intelligent defense system capable of identifying attacks and safeguarding infrastructures against increasingly sophisticated cyber threats.

METRICS FOR EVALUATING ANOMALY DETECTION IN AI-BASED CYBER DEFENSE

In the context of intelligent defense against cyber threats, AI-based anomaly detection systems are evaluated using metrics such as accuracy, precision, recall - detection rate, false positive rate (FPR), detection latency, and area under the curve (AUC). Anomaly detection models must be measured not only by their ability to identify malicious activity but also by their reliability, efficiency, and operational feasibility. The metrics, with descriptions and references about their use, are presented in Table 2.

Table 2: *Summary of key metrics for AI anomaly detection in cyber defense*

Metric	Description	Typical values in recent studies	Citations
Accuracy	Proportion of correctly identified anomalies (true positives + true negatives) among all cases	94% – 99.98%	(Bondarenko & Statsenko, 2024; Chaudhary et al., 2024; Khalaf et al., 2024; Kalutharage, Liu & Chrysoulas, 2025; Reddy &Vani, 2025)
Precision	Proportion of detected anomalies that are actual threats (true positives / all positives)	93% – 99.4%	(Chaudhary et al., 2024; Dorothy et al., 2024; Kafita & Yamazaki, 2024; Reddy &Vani, 2025)
Recall	Proportion of actual threats correctly detected (true positives / all actual threats)	90% – 98.8%	(Chaudhary et al., 2024; Dorothy et al., 2024; Kafita & Yamazaki, 2024; Abdulrahman et al., 2025; Reddy &Vani, 2025)
FPR	Proportion of benign events incorrectly flagged as threats	1.1% – 2.0%	(Chaudhary et al., 2024; Reddy &Vani, 2025)
Detection Latency	Time taken to identify an anomaly after it occurs	As low as 45ms	(Reddy & Vani, 2025)
AUC	Overall ability to distinguish between threats and benign events	Up to 0.96	(Chaudhary et al., 2024; Kafita & Yamazaki, 2024)

Additional considerations are:

- Explainability and Interpretability: Modern systems increasingly use explainable AI (XAI) to clarify why an anomaly is flagged, improving trust and response speed (Moustafa et al., 2023; Kalutharage, Liu & Chrysoulas, 2025);
- Adaptability: AI models are evaluated on their ability to detect novel or zero-

day attacks, not just known threats (Bondarenko & Statsenko, 2024; Kalutharage, Liu & Chrysoulas, 2025 Reddy & Vani, 2025);

- Resource Efficiency: Feature selection and model optimization are used to maintain high detection rates with minimal computational overhead (Kafita & Yamazaki, 2024; Kalutharage, Liu & Chrysoulas, 2025).

Comparing these metrics shows that modern AI-driven anomaly detection systems deliver high accuracy, precision, and recall, while minimizing false positives and detection latency. This makes them highly effective for intelligent cyber defense, especially in dynamic and high-risk environments.

AI-driven anomaly detection in cyber defense is measured by a combination of accuracy, precision, recall, FPR, and detection latency. Recent research demonstrates that advanced AI models can achieve high detection rates (often above 95%) with low false positives and rapid response times, making them highly effective for modern cybersecurity needs.

The most relevant performance metrics, along with their mathematical formulations and contextual significance in cybersecurity applications, are presented below.

The True Positive Rate (TPR), often referred to as Recall or Sensitivity, measures the proportion of actual anomalies that are correctly identified by the system. It reflects the system's ability to capture malicious behavior and is significant in security environments where undetected threats can lead to severe breaches. The calculus formula is:

$$TPR = \frac{TP}{TP + FN}$$

where: *TPR* =the true positive rate; *TP* (True Positives) = anomalies correctly detected; *FN* (False Negatives) = anomalies that remain undetected

A higher *TPR* indicates that fewer attacks slip through the defense system. However, models that maximize recall may also generate more false alarms, which necessitates a balance with other metrics.

The False Positive Rate (FPR) quantifies the proportion of benign activities incorrectly labeled as anomalies:

$$FPR = \frac{FP}{FP + TN}$$

where: FPR=the False Positive Rate: FP (False Positives) = normal instances flagged as anomalies; TN (True Negatives) = normal instances correctly identified.

In cybersecurity operations, a high FPR can overwhelm analysts with spurious alerts, leading to "alert fatigue". Consequently, maintaining a low FPR is as critical as maximizing TPR.

Precision evaluates the proportion of detected anomalies that are truly malicious:

$$Precision = \frac{TP}{TP + FP}$$

High precision ensures that analysts can trust the alerts they receive. However, optimizing only for precision risks neglects recall, as the system might only report anomalies when it is certain, leaving some threats undetected.

Accuracy represents the overall proportion of correct predictions across all categories (both normal and anomalous) and is calculated by the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

While accuracy is intuitive, it can be misleading in imbalanced datasets, which are common in cybersecurity. A system that labels all events as normal may achieve high accuracy, yet completely fail at threat detection.

F1-Score. To strike a balance between Precision and Recall, the *F1-Score* is employed:

$$F1 - Score = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

The *F1-Score* is particularly valuable in anomaly detection because it penalizes systems that favor one metric over the other.

ROC-AUC (Receiver Operating Characteristic – Area Under Curve) provides a threshold-independent

measure of the system's ability to discriminate between normal and anomalous instances.

PR-AUC (Precision-Recall Area Under Curve) is more informative in highly imbalanced datasets, typical of network traffic. While *ROC-AUC* may appear artificially high in such settings, *PR-AUC* highlights the trade-offs between the model's credibility and completeness.

Detection Latency. In real-time cyber defense, this is a crucial metric. It measures the average time delay between the occurrence of an anomaly and its detection.

The calculus formula is:

$$Latency = \frac{\sum(t_{detection} - t_{event})}{N_{anomalies}}$$

where: C_{FN}=relative costs of false negatives; C_{FP}=relative costs of false positives.

In cybersecurity, models are usually optimized to minimize the probability of costly false negatives while maintaining manageable levels of false positives.

where: $t_{detection}$=timestamp when anomaly is detected; t_{event}=timestamp when anomaly occurred; $N_{anomalies}$=number of true anomalies.

Low latency is vital for preventing escalation of attacks such as ransomware, where even a few minutes of delay can lead to irreparable damage.

Cost-Based Metrics. Organizations often adopt cost-sensitive evaluations to balance operational risks. A *False Negative* (FN) (missed attack) typically incurs much higher costs than a *False Positive* (FP) (false alarm). The cost is calculated:

$$Cost = C_{FN} * FN + C_{FP} * FP$$

Recent research on AI-based anomaly detection across various domains (including cybersecurity, networks, IoT, and video surveillance) reports detection rates (recall) typically ranging from 86% to nearly 100%, with many systems also achieving high accuracy and low false positive rates (Table 3).

Table 3: *Actual anomaly detection rates and related metrics from recent research*

Application Domain	Detection Rate / Recall	Accuracy	False Positive Rate	Citations
Network traffic (Genetic Algorithm + Fuzzy Logic)	96.53%	96.53%	0.56%	(Hamamoto et al., 2018)
Wireless sensor networks (OLWPR)	86%	Not reported	16% error rate	(Poornima & Paramasivan, 2020)
Streaming network data (Evolving Clustering)	Nearly 100%	98.86%	1.25%	(Wang et al., 2022)
Industrial cyber-physical systems (Autoencoder)	87.5% – 96.5%	87.5% – 96.5%	Not reported	(Du et al., 2024)
Video anomaly detection (Spatio-temporal)	98.4% (AUC)	Not reported	Not reported	(Wang et al., 2023)
Building energy (Autoencoder)	94.5%	Not reported	4.7%	(Araya et al., 2017)
IoT devices (Game-theoretic IDS)	93%	Not reported	2%	(Sedjelmaci, Senouci & Taleb, 2017)
Indoor air quality (LSTM-Autoencoder)	>99%	99.5%	Not reported	(Wei et al., 2022)

High-performing AI models (e.g., DL, autoencoders, clustering) consistently achieve detection rates above 90%, with some reporting near-perfect results in controlled or benchmark datasets (Hamamoto et al., 2018; Wang et al., 2022; Wei et al., 2022). False positive rates are generally low (often below 5%), which is crucial for practical deployment (Araya et al., 2017; Sedjelmaci, Senouci & Taleb, 2017; Hamamoto et al., 2018; Wang et al., 2022). Detection rates can vary by application, data quality, and method, with some domains (e.g., wireless sensor networks) reporting lower rates due to environmental complexity (Poornima & Paramasivan, 2020).

State-of-the-art AI anomaly detection systems typically achieve detection rates between 86% and 100%, with many reporting high accuracy and low false positive rates. These results demonstrate the strong potential of AI for intelligent cyber defense and other anomaly detection applications.

The evaluation of AI-driven anomaly detection in cyber defense requires a multi-metric approach. While Recall emphasizes the importance of not missing attacks, Precision and FPR address the operational burden of false alarms. Advanced measures such as ROC-AUC, PR-AUC, and detection latency provide deeper insights into real-world applicability, while cost-based metrics align evaluation with organizational priorities. Together, these metrics ensure that AI systems are not only technically sound but also operationally effective in strengthening intelligent defense strategies.

ARTIFICIAL INTELLIGENCE TECHNOLOGIES AND CYBERSECURITY SOLUTIONS

AI technologies in cybersecurity are defined by their capacity to autonomously analyze, learn from large volumes of data, and continuously adapt to new attack patterns and threat landscapes (Khan, M.I., Arif & Khan, A.R.A., 2024). Unlike traditional rule-based systems, AI employs ML models, DL neural networks, and Natural Language Processing (NLP) to detect threats, identify anomalies, predict attacks, and automate responses. This targeted application of AI, which combines data-driven adaptability with automation, addresses complex digital security challenges.

In the face of increasingly frequent and various threats that bypass traditional defenses, cybersecurity must adopt hybrid approaches that combine rules-based methods, AI, and ML to optimize threat response. Integrating AI into security infrastructures enhances threat alert sorting, prioritizes detection, classifies vulnerabilities, automates patching, accelerates data processing, and improves configuration management. The most recognized AI-based technologies and security solutions are detailed below.

Machine Learning and Deep Learning

ML and DL form the backbone of AI in cybersecurity, utilizing algorithms that learn from data to identify threats and patterns in cyber activity. DL, an advanced form of ML that uses deep neural networks, is used for complex tasks like malware classification and intrusion detection. The key applications for ML and DL are numerous. *Intrusion Detection/ Prevention Systems* (IDS/IPS), based on ML and DL algorithms, analyze large volumes of data in real-time, identifying abnormal behaviors and automating the analysis of datasets (Fattahi, 2024a; Xu, 2025). These systems can detect previously unknown zero-day attacks and novel malware by analyzing behavioral characteristics and abnormal traffic patterns (Das & Rao, 2025); widely used datasets for training include UNSW-NB15 and CICIDS 2017/2018. Furthermore, User and Entity Behavior Analytics (UEBA) utilizes AI to continuously monitor user and device activity to establish a behavior baseline, allowing DL models to identify deviations that signal compromised accounts or internal threats, such as unusual access times or locations (Wang et al., 2023). Lastly, *Malware and Phishing Detection* heavily relies on DL models, such as CNNs, to rapidly analyze file binary features and web page images to identify and block threats.

Natural Language Processing

NLP enables AI systems to understand, interpret, and generate human language. Its key applications include generating threat intelligence, where NLP automatically extracts valuable information about threats (malicious IPs, tactics, techniques, and procedures) from unstructured sources like security reports, hacking forums, or the dark web. It is also critical for Phishing and Spam Detection, as NLP models analyze the subject, body, and tone of an email to identify social engineering attempts that mimic legitimate communications.

Generative AI

GenAI uses large models, such as Large Language Models (LLMs) or Generative Adversarial Networks (GANs), to create new content. Its key applications in Offensive Security (Red Team) involve accelerating the creation of new malicious code or the personalization of sophisticated phishing attacks, thereby facilitating penetration testing and vulnerability identification (Sarker et al., 2024b). Conversely, for Defensive Security (Blue Team), GenAI can generate synthetic data for training detection models, enhancing their effectiveness without compromising real data, and assisting security analysts in summarizing complex incidents and alerts.

Expert Systems and Explainable AI

Expert Systems (ES) utilize codified knowledge based on rules and logical reasoning, while XAI addresses the model opacity problem, making AI decisions transparent and justifiable. These technologies are important for Security Orchestration, Automation, and Response (SOAR). Platforms utilize ES and AI to orchestrate and automate incident response tasks, such as isolating compromised hosts or running malware analysis, drastically reducing response time. In Fraud Detection and Prevention, particularly in financial transactions, AI analyzes transactional connections in real-time to detect illegitimate patterns, relying on ML models that can be enhanced with XAI reasoning for auditability. Lastly, in Risk Management, AI identifies and assesses risks in large datasets through automated, real-time analysis, detecting vulnerabilities and suspicious behaviors. Tools like IBM QRadar Advisor with Watson and Microsoft Sentinel perform predictive analytics and continuous risk assessment, thereby optimizing security controls.

Summary of AI Technologies in Cybersecurity

Table 4 presents a synthesis of the main AI technologies and their specific applications in cybersecurity.

Table 4: *AI Technologies Used in Cybersecurity*

Technology	Brief Description	Key Applications in Cybersecurity	References
ML	Algorithms that learn from data to detect threats, anomalies, and patterns in cyber activity	Intrusion detection, malware classification, anomaly detection, phishing detection, behavioral analysis, and threat intelligence	(Kaur, Gabrijelčič & Klobučar, 2023; Dhanushkodi & Thejas, 2024; Ali, Wang & Leung, 2025)
DL	Advanced ML using neural networks for complex tasks like malware classification and intrusion detection	Malware detection, network traffic analysis, vulnerability search, intrusion detection, source localization, and steganalysis	(Sarker, Furhad & Nowrozy, 2021; Mohamed, 2023; Ali, Wang & Leung, 2025; Siam et al., 2025)

NLP	AI that understands and analyzes human language for threat intelligence, phishing detection, and vulnerability analysis	Phishing email detection, threat intelligence extraction, vulnerability analysis, static code analysis, and malware detection	(Sarker, Furhad & Nowrozy, 2021; Dhanushkodi & Thejas, 2024; Ali, Wang & Leung, 2025)
Expert Systems and Decision	Rule-based AI and multi-criteria decision-making that automates decision-making for threat response and security policy enforcement	Automated threat response, security policy enforcement, context reasoning, and access authentication	(Sarker, Furhad & Nowrozy, 2021; Zhang et al., 2021, Rădulescu et al., 2025)
GenAI	AI that creates new content (e.g., text, code), used for both cyber defense and attacks	Automated code review, threat simulation, phishing and social engineering, malware creation, and attack payload generation	(Gupta et al., 2023)
XAI	AI systems are designed to provide transparent, understandable reasoning for security decisions and threat detection	Explainable intrusion detection, malware detection, phishing/ spam detection, digital forensics, and fraud detection	(Capuano et al., 2022; Sarker et al., 2024a)
Blockchain-integrated AI	Combines AI with blockchain for decentralized, tamper-resistant, and privacy-preserving cybersecurity solutions	Secure data sharing, decentralized threat intelligence, privacy-preserving AI, and tamper-resistant security logs	(Saleh, 2024; Ali, Wang & Leung, 2025)
MFA	Requires two or more independent credentials for user authentication, increasing security over single-factor methods	Secure access to systems, protection of digital identities, transaction security, biometric and token-based authentication, and healthcare data access	(Ahmad et al., 2023; Rahman, Titouna & Naït-Abdesselam, 2024; Rivera, Muhammad & Song, 2024)

AI technologies (especially ML and DL) are increasingly used in cybersecurity, especially for detection, automation, threat intelligence, and incident response. They bring advantages: scale, speed, ability to handle novel threats, but also challenges: adversarial attack, data quality, explainability, and resource constraints. The frontier seems to be hybrid reasoning-plus-learning systems, strengthening defensive AI, and embedding ethics and governance into deployments.

CHALLENGES AND LIMITATIONS

Although AI brings remarkable advances in detecting and responding to cyber threats, its widespread application involves several technical, operational, and ethical challenges that cannot be ignored.

Understanding these limitations is essential for developing robust, secure, and responsible systems.

False positives and false negatives

One of the most significant problems in AI applications for cybersecurity is the occurrence of an increased rate of false positives and false negatives.

False positives occur when legitimate behavior is labeled as malicious, which can lead to unnecessary service interruptions and alert fatigue. False negatives allow real attacks to go unnoticed, compromising system security.

This imbalance is exacerbated in dynamic environments or when faced with unknown (zero-day) attacks, where AI models, especially those based on supervised learning, struggle to generalize correctly beyond the training set.

Model opacity and lack of explainability

Advanced DL models, such as CNN or LSTM networks, are often considered "black boxes" due to their internal complexity. This makes it difficult to explain the decisions made and affects organizations' trust in the results of AI systems, representing an obstacle to widespread adoption, especially in regulated areas such as financial security or critical infrastructure.

Modern approaches attempt to remedy this by integrating XAI components that provide visualizations, importance scores, and arguments for the decisions made. However, the level of interpretability remains limited for deep models or unstructured data.

Adversarial attacks (Adversarial AI)

AI systems are vulnerable to adversarial attacks, in which inputs are subtly modified to fool the model. For example, a network file or email can be altered so that malware is interpreted as benign or a phishing message goes undetected. These attacks highlight the architectural weaknesses of AI, including overreliance on patterns in training data and a lack of robustness to perturbations. In some cases, adversaries even use AI models to generate these malicious inputs (offensive AI), creating a cycle of AI vs. AI.

Expertise gaps

The shortage of qualified professionals makes it difficult to implement, update, and manage essential protection measures. Without proper training, errors and vulnerabilities are more likely, which slows incident response and weakens overall system security.

Resource demands

Effective cybersecurity depends not only on robust infrastructure and advanced monitoring tools, but also on well-trained staff, clear incident response plans, and strict compliance with regulations. Equally important is the ability to encrypt and safeguard data, maintain backup systems, and collaborate closely with authorities and suppliers to ensure continuity and trust in digital services.

Ethical issues and lack of accountability

As AI assumes increasingly sensitive tasks in cyber defense (e.g., automatic traffic blocking, isolation of compromised systems, reporting to authorities), ethical concerns arise regarding decision-making autonomy and the lack of human accountability. Such systems may affect users or critical infrastructures without oversight, raising worries about traceability, responsibility in the event of errors, and biases stemming from historical training data (Fattahi, 2024b). Risks of excessive surveillance or algorithmic discrimination also arise if predictive analytics lack transparency and auditability (Neupane et al., 2022; Xu et al., 2025). For AI systems to earn trust, they must assist human decision-making, be technically robust and safe, protect data privacy, and remain transparent and explainable. Equally important are fairness, accountability, and awareness of social and environmental impacts, ensuring that AI serves people responsibly and predictably.

EXAMPLES OF AI-BASED PLATFORMS IN REAL-WORLD APPLICATIONS

The adoption of AI in cybersecurity is no longer a promise of the future, but an active reality within global organizations. Various commercial solutions and enterprise platforms employ AI to detect, prevent, and respond to sophisticated attacks. The following examples provide insight into the effectiveness and diversity of AI applications in real-world environments.

Darktrace is one of the most well-known commercial platforms that leverages self-learning AI to identify insider threats, such as unauthorized access or unusual employee behavior. The system employs unsupervised learning algorithms to autonomously establish a baseline of an organization's "digital normal" and subsequently detect deviations from this model, without relying on predefined rules. In a case reported by Darktrace, AI uncovered the misuse of multiple user accounts, including ones with elevated privileges, in a SaaS environment. The compromised accounts were accessed from unusual locations and used to create malicious inbox rules. This subtle abuse bypassed traditional security measures, but Darktrace's anomaly detection identified the suspicious behavior and prevented further data loss.

Google Chronicle Security, developed by Alphabet's former X incubator, is a platform designed for the large-scale analysis of security logs using AI. It leverages scalable storage and AI models to correlate suspicious events across enterprise IT infrastructures in real time. A use case involved the early-stage detection of ransomware attacks, where abnormal file behaviors and Server Message Block (SMB) transfers were associated with known attack patterns (via ML), enabling the preventive isolation of infected systems.

IBM Safer Payments is a configurable platform for preventing fraud in cashless payments, through all methods practiced by banks, with which fraudulent transactions are detected and blocked. Customer transactions are analyzed and assessed in a risk score; fewer transactions are blocked or unnecessarily checked, and the false positive rate is very low. According to IBM, one implementation example involved the interbank company Stet, which reported a decrease in fraud losses, thus achieving benefits of over $100 million annually.

The examples analyzed here clearly demonstrate that AI is no longer merely an experimental tool in the field of cybersecurity but rather an essential component of the defensive architecture of modern organizations. From identifying anomalous user behavior to preventing the automated spread of malware and detecting advanced phishing attempts generated by large language models, AI has proven capable of ensuring not only rapid detection but also efficient real-time response. Platforms such as Darktrace, Google Chronicle, and IBM Safer Payments reflect the growing maturity of the market in terms of AI integration into commercial security systems, showing that these technologies can be applied at scale with measurable results. Moreover, the adoption of AI in security operations centers enables a significant reduction in the response time and facilitates the intelligent prioritization of alerts, thereby reducing the risk of alert fatigue among analysts (Markevych & Dawson, 2023).

FUTURE DIRECTIONS IN INTELLIGENT CYBER DEFENSE

Intelligent cyber-defense is rapidly evolving to address the increasing sophistication, scale, and automation of cyber threats. The integration of AI, ML, DL, and advanced analytics is transforming how organizations detect, prevent, and respond to cyberattacks. Recent literature highlights a shift from reactive to proactive, adaptive, and autonomous defense strategies, with a focus on explainability, resilience, and ethical considerations. Key future directions include the development of autonomous cyber-defense agents, proactive threat intelligence, integration with emerging technologies (such as blockchain and quantum computing), and the advancement of explainable and trustworthy AI systems.

Autonomous and adaptive cyber defense agents

The future of cyber-defense is moving toward autonomous intelligent agents capable of real-time detection, response, and even counterattacks against AI-driven threats. These agents, such as Autonomous Intelligent Cyber-defense Agents (AICAs), are being developed for military, critical infrastructure, and enterprise environments, with a focus on scalability, adaptability, and minimal human intervention ((Blakely, 2022; Oesch et al., 2024; Holz, Loevenich, & Lopes 2025).

Proactive and predictive security

There is a strong trend toward proactive defense mechanisms, including AI-driven predictive analytics, cyber threat intelligence mining, and moving target defense. These approaches aim to anticipate and prevent attacks before they occur, leveraging real-time data, behavioural analytics, and continuous feedback loops for model retraining and adaptation (Ali, Wang & Leung, 2025; Hasan et al., 2025; Reddy & Vani, 2025).

Explainable, trustworthy, and ethical AI

As AI systems become more integral to cyber defense, explainability, transparency, and ethical governance are critical. XAI is being developed to provide interpretable outputs, foster trust, and support regulatory compliance. Research also emphasizes the need for robust ethical frameworks and international cooperation to address the risks of algorithmic opacity and adversarial misuse (Agarwal, 2025; Sarker et al., 2024b).

Integration with emerging technologies

The integration of AI with blockchain, quantum computing, edge computing, and big data analytics is a key research direction. Blockchain enhances data integrity and decentralized authentication, quantum computing introduces new paradigms for both attack and defense, and edge computing enables real-time, localized threat mitigation. These technologies collectively support the development of resilient, scalable, and future-proof cyber-defense systems (Ali, Wang & Leung, 2025; Hasan et al., 2025).

The literature demonstrated a clear trajectory toward autonomous, adaptive, and proactive cyber-defense systems powered by advanced AI and hybrid models. However, challenges remain, including adversarial attacks on AI models, data privacy, explainability, and the integration of emerging technologies. The need for standardized metrics, robust datasets, and interdisciplinary collaboration is emphasized across the literature (Kaur, Gabrijelčič & Klobučar, 2023; Ali, Wang & Leung, 2025; Hasan et al., 2025).

CONCLUSIONS

The paper highlights that the use of AI in the field of cybersecurity has become a strategic necessity rather than merely a technological option. It emphasizes the capacity of AI to contribute to the development of an intelligent defense model capable of addressing increasingly sophisticated cyberattacks. It proves that AI provides significant advantages over traditional solutions: rapid identification of suspicious patterns, anticipation of attacks, and near real-time response. Thus, AI can provide particular relevant solutions to protect critical infrastructures, in different areas such as healthcare, the energy sector, and national security, where a cyberattack can have severe consequences with social, economic, and political impact.

At the same time, the paper underlines the limitations of implementing these technologies. The adoption of AI requires considerable resources, specialized expertise, and heightened attention to the ethical implications of automated decision-making. These aspects emphasize the importance of developing an appropriate regulatory framework, both nationally and internationally, to ensure a balance between innovation, efficiency, and responsibility.

AI represents both a protective shield and a strategic challenge in cybersecurity. Future research directions should focus on

refining defense algorithms, building control mechanisms, transparency, and accountability, and also integrating with emerging technologies. In this way, authentic resilience can be achieved, capable of protecting society against ever-evolving cyber risks.

ACKNOWLEDGEMENTS

This work was carried out through the Core Program within the National Research Development and Innovation Plan 2022-2027, carried out with the support of MCID, project no. 23380101, „Contributions to the consolidation of emerging technologies specific to the Internet of Things and complex systems".

REFERENCE LIST

Abdulrahman, I.A., Ogor, U.C., Ayodele, G.T., Anadozie, C. & Alebiosu, J. (2025) AI-Driven Threat Intelligence and Automated Incident Response: Enhancing Cyber Resilience through Predictive Analytics. *Research Journal in Civil, Industrial and Mechanical Engineering.* 2(1), 16-32. https://doi.org/10.61424/rjcime.v2i1.236.

Adi, E., Baig, Z. & Zeadally, S. (2022) Artificial Intelligence for Cybersecurity: Offensive Tactics, Mitigation Techniques and Future Directions. *Applied Cybersecurity and Internet Governance.* 1(1), 1–23. https://doi.org/10.5604/01.3001.0016.0800.

Agarwal, G. (2025) Explainable AI (XAI) for Cyber Defense: Enhancing Transparency and Trust in AI-Driven Security Solutions. *International Journal of Advanced Research in Science, Communication and Technology.* 5(1), 132–138. https://doi.org/10.48175/IJARSCT-23624.

Ahmad, M.O., Tripathi, G., Siddiqui, F., Alam, M.A., Ahad, M.A., Akhtar, M.M. & Casalino, G. (2023) BAuth-ZKP—A Blockchain-Based Multi-Factor Authentication Mechanism for Securing Smart Cities. *Sensors.* 23(5), 2757. https://doi.org/10.3390/s23052757.

Ainslie, S., Thompson, D., Maynard, S. & Ahmad, A. (2023) Cyber-threat intelligence for security decision-making: A review and research agenda for practice. *Computers & Security.* 132, 103352. https://doi.org/10.1016/j.cose.2023.103352

Ali, S., Wang, J. & Leung, V.C.M. (2025) AI-driven fusion with cybersecurity: Exploring current trends, advanced techniques, future directions, and policy implications for evolving paradigms—A comprehensive review. Information Fusion. 118, 102922. doi:10.1016/j.inffus.2024.102922.

Araya, D.B., Grolinger, K., ElYamany, H.F., Capretz, M.A.M. & Bitsuamlak, G. (2017) An ensemble learning framework for anomaly detection in building energy consumption. *Energy and Buildings.* 144, 191-206. https://doi.org/10.1016/J.ENBUILD.2017.02.058.

Blakely, B. (2022) An Experimental Platform for Autonomous Intelligent Cyber-Defense Agents: Towards a collaborative community approach (WIPP). *2022 Resilience Week (RWS), 26-29 September 2022, National Harbor, MD, USA, IEEE.* pp. 1-7. https://doi.org/10.1109/RWS55399.2022.9984037.

Bondarenko, A. & Statsenko, V. (2024) Use of Artificial Intelligence methods and models for improving expert systems of intrusion detection. Herald of Khmelnytskyi National University. *Technical sciences.* 333 (2), 99-106. https://doi.org/10.31891/2307-5732-2024-333-2-15.

Capuano, N., Fenza, G., Loia, V. & Stanzione, C. (2022) Explainable Artificial Intelligence in CyberSecurity: A Survey. *IEEE Access.* 10, 93575–93600. https://doi.org/10.1109/ACCESS.2022.3204171.

Chaudhary, D., Verma, S.K., Shrimal, V. M., Madala, R., Baliyan, R. & Satish, M. (2024). AI-Based Methods to Detect and Counter Cyber Threats in Cloud Environments to Strengthen Cloud Security. *2024 International Conference*

on Electrical Electronics and Computing Technologies (ICEECT), 29-31 August 2024, Greater Noida, India, IEEE. pp. 1-6. https://doi.org/10.1109/ICEECT61758.2024.10739173.

Das, R. & Rao, G.N. (2025) Intrusion Detection System for Network Attacks. *International Journal for Research in Applied Science & Engineering Technology.* 13(7), 2395–2402. https://doi.org/10.22214/ijraset.2025.73398.

Dhanushkodi, K. & Thejas, S. (2024) AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation. *IEEE Access.* 12, 173127–173136. https://doi.org/ 10.1109/ACCESS.2024.3493957.

Dorothy, A.B., Madhavidevi, B., Nachiappan, B., Manikandan, G., Patjoshi, P.K. & Sindhuja, M. (2024) AI-Driven Threat Intelligence in Cloud Computing: Detecting and Responding to Cyber Attacks. *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), 23-24 August 2024, Hassan, India, IEEE.*pp.1-6. https://doi.org/10.1109/IACIS61494. 2024.10721888.

Du, X., Zhou, C.-Y, Tian, Y. & Wang, K. (2024) Anomaly Detection Based on Data Super-Resolution in Industrial Cyber–Physical Systems with Multirate Sampling. *IEEE Sensors Journal.* 24(10), 16478-16490. https://doi.org/10.1109/JSEN.2024.3384275.

Fattahi, J., Machine Learning and Deep Learning Techniques used in Cybersecurity and Digital Forensics: a Review. To be published in Cryptography and Security. [Preprint] https://arxiv.org/abs/2501.03250 [Accessed: 9th December 2025]

Fattahi, J., (2024b). Machine Learning and Deep Learning Techniques used in Cybersecurity and Digital Forensics: a Review. To be published in *Cryptography and Security.* [Preprint] https://arxiv.org/abs/2501.03250 [Accessed: 9th December 2025]

Ferrag, M. A., Alwahedi, F., Battah, A., Cherif, B., Mechri, A., Tihanyi, N., Bisztray, T., & Debbah, M. (2025). Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities. *Internet of Things and Cyber-Physical Systems*. 5, 1-46. https://doi.org/10.1016/j.iotcps.2025.01.001.

Franco, J., Aris, A., Canberk, B., & Uluagac, A. S. (2021). A survey of honeypots and honeynets for Internet Of Things, Industrial Internet of Things, and Cyber-Physical Systems. *IEEE Communications Surveys & Tutorials*. 23(4), 2351–2383. https://doi.org/10.1109/COMST.2021.3106669.

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access*. 11, 80218–80245. https://doi.org/10.1109/ACCESS.2023.1234567.

Hamamoto, A.H., Carvalho, L.F., Sampaio, L.D.H., Abrão, T., & Proença Jr., M.L. (2018) Network Anomaly Detection System using Genetic Algorithm and Fuzzy Logic. *Expert Systems with Applications.* 92, 390-402. https://doi.org/10.1016/j.eswa.2017.09.013.

Hasan, K., Hossain, F., Amin, A., Sutradhar, Y., Jeny, I. J. & Mahmud, S. (2025) Enhancing Proactive Cyber Defense: A Theoretical Framework for AI-Driven Predictive Cyber Threat Intelligence. *Journal of Technologies Information and Communication.* 5(1), 33122. https://doi.org/10.55267/rtic/16176.

Holz, L., Loevenich, J. & Lopes, R.R.F. (2025) Towards Robust Autonomous Cyber Defence Agents Using Hybrid AI Models. *Proceedings of the 2025 IEEE 11th International Conference on Network Softwarization (NetSoft), 23-27 June 2025, Budapest, Hungary, IEEE.* pp. 269–272. https://doi.org/10.1109/NetSoft64993.2025.11080605

Kafita, E.-V & Yamazaki, T. (2024) Leveraging Artificial Intelligence Feature Selection for Cybersecurity Network Anomaly Intrusion Detection. *2024 7th World Symposium on Communication Engineering (WSCE), 28-30 September 2024, Tokyo, Japan, IEEE.* pp.22-26. https://doi.org/10.1109/WSCE65107.2024.00010.

Kalutharage, C., Liu, X. & Chrysoulas, C. (2025) Neurosymbolic learning and domain knowledge-driven explainable AI for enhanced IoT network attack detection and response. *Computer & Security.* 151, 104318. https://doi.org/10.1016/j.cose.2025.104318.

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

Khalaf, L., Alhamadani, B., Ismael, O., Radhi, A., Ahmed, S. & Algburi, S. (2024) Deep Learning-Based Anomaly Detection in Network Traffic for Cyber Threat Identification. *AICCONF '24: Proceedings of the Cognitive Models and Artificial Intelligence Conference.* Association for Computing Machinery, New York, NY, United States.

Khan, M.I., Arif, A., & Khan, A.R.A. (2024) The Most Recent Advances and Uses Of AI in Cybersecurity. *BULLET: Jurnal Multidisiplin Ilmu.* 3(4), 566–578.

Kumari, I., & Lee, M. (2023). A prospective approach to detect advanced persistent threats: Utilizing hybrid optimization technique. *Heliyon.* 9(11), e21377. https://doi.org/10.1016/j.heliyon.2023.e21377.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature.* 521(7553), 436–444. https://doi.org/10.1038/nature14539.

Markevych, M. & Dawson, M. (2023) A Review of Enhancing Intrusion Detection Systems for Cybersecurity Using Artificial Intelligence (AI). *International Conference Knowledge-Based Organization.* 29(3), 30–37. https://doi.org/10.2478/kbo-2023-0005.

Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering.*

10(2), 2272358. https://doi.org/10.1080/23311916.2023.2272358.

Moustafa, N., Koroniotis, N., Keshk, M., Zomaya, A. & Tari, Z. (2023) Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions. *IEEE Communications Surveys & Tutorials.* 25(3), 1775-1807. https://doi.org/10.1109/COMST.2023.3280465.

Neupane, S., Ables, J., Anderson, W., Mittal, S., Rahimi, S., Bănicescu, I. & Seale, M. (2022) Explainable Intrusion Detection Systems (X-IDS): A Survey of Current Methods, Challenges, and Opportunities. *IEEE Access.* 10 (7), 112392-112415. https://doi.org/10.1109/ACCESS.2022.3212345.

Oesch, S., Austria, P., Chaulagain, A., Weber, B., Watson, C., Dixson, M. & Sadovnik, A. (2024) The Path to Autonomous Cyberdefense. *IEEE Security & Privac.* 23, 38–46. https://doi.org/10.1109/MSP.2024.1234567.

Poornima, I. & Paramasivan, B. (2020) Anomaly detection in wireless sensor network using machine learning algorithm. *Computer Communications.* 151, 331-337. https://doi.org/10.1016/j.comcom.2020.01.005.

Rahman, F., Titouna, C. & Naït-Abdesselam, F. (2024) Security Enhancement Using Scalable Blockchain-Based Multi-Factor Authentication. In *Proceedings of the 2024 6th International Conference on Blockchain Computing and Applications (BCCA).* 428–433. IEEE. https://doi.org/10.1109/BCCA.2024.1234567.

Rădulescu, C.Z., Rădulescu, M., Vevera, A.-V. & Boncea, R. (2025) A Hybrid Approach Based on the Two Weight Vectors and Extended TOPSIS Methods With Application in Cybersecurity. *International Journal of Information Technology & Decision Making.* https://doi.org/10.1142/S0219622025501135. [Accessed: 8th January 2025]

Reddy, S., Chaudhari, T., Godla, S., Venkata, J., Ramesh, N., Muniyandy, E., Kranthi, A., & El-Ebiary, D. (2025). AI-Driven Transformer Frameworks for Real-Time Anomaly Detection in Network Systems. *International Journal of Advanced Computer Science and Applications.* https://doi.org/10.14569/ijacsa.2025.01602111.

Reddy, P.M. & Vani, S.M. (2025) Cyber Threat Intelligence Analysis for Proactive Cybersecurity Defense: A Survey and New Perspectives. *Journal of Engineering Sciences.* 25(3), 1748–1774. https://doi.org/10.36893/jes.2025.v16i04.018.

Rivera, J., Muhammad, A., & Song, W. (2024). Securing digital identity in the zero-trust architecture: A blockchain approach to privacy-focused multi-factor authentication. *IEEE Open Journal of the Communications Society.* 5, 2792–2814. https://doi.org/10.1109/OJCOMS.2024.1234567.

Saleh, A.M.S. (2024) Blockchain for secure and decentralized AI in cybersecurity: A comprehensive review. *Blockchain: Research and Applications.* 3, 100193. https://doi.org/10.1016/j.bcra.2024.100193.

Sarker, I.H., Furhad, M.H. & Nowrozy, R. (2021) AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling, and Research Directions. *SN Computer Science.* 2, 173. https://doi.org/10.1007/s42979-021-00557-0.

Sarker, I.H., Janicke, H., Mohsin, A., Gill, A. & Maglaras, L. (2024a) Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges, and prospects. ICT Express. 10(4), 935–958. https://doi.org/10.1016/j.icte.2024.05.007.

Sarker, I.H., Janicke, H., Ferrag, M.A., & Abuadbba, A. (2024b). Multi-aspect rule-based AI: Methods, taxonomy, challenges and directions towards automation, intelligence and transparent cybersecurity modeling for critical infrastructures. *Internet of Things.* 25, 101110. https://doi.org/10.1016/j.iot.2024.101110.

Sedjelmaci, H., Senouci, S.M. & Taleb, T. (2017) An Accurate Security Game for Low-Resource IoT Devices. *IEEE Transactions on Vehicular Technology.* 66 (10), 9381-9393. https://doi.org/10.1109/TVT.2017.2701551.

Siam, A., Alazab, M., Awajan, A., & Faruqui, N. (2025). A comprehensive review of AI's current impact and future prospects in cybersecurity. IEEE Access. 13, 14029–14050. https://doi.org/10.1109/ACCESS.2025.1234567.

Sutton, R.S. & Barto, A.G. (1998) Reinforcement learning: An introduction. *IEEE Transactions on Neural Networks.* 9(6), 1054–1054. https://doi.org/10.1109/72.711337.

Wang, X., Ahmed, M.M., Husen, M.N., Qian, Z. & Belhaouari, S.B. (2022) Evolving anomaly detection for network streaming data. *Information Sciences.* 608, 757-777. https://doi.org/10.1016/j.ins.2022.06.064.

Wang, Y., Liu, T., Zhou, J. & Guan, J. (2023) Video anomaly detection based on spatio-temporal relationships among objects. *Neurocomputing.* 532, 141-151. https://doi.org/10.1016/j.neucom.2023.02.027.

Wei, Y., Jang-Jaccard, J., Xu, W., Sabrina, F., Çamtepe, S. & Boulic, M. (2022) LSTM-Autoencoder-Based Anomaly Detection for Indoor Air Quality Time-Series Data. *IEEE Sensors Journal.* 23(4), 3787-3800. https://doi.org/10.1109/JSEN.2022.3230361.

Xu, Z., Wu, Y., Wang, S., Gao, J., Qiu, T., Wang, Z., Wan, H. & Zhao, X. (2025) Deep learning-based intrusion detection systems: A survey. To be published in *Cryptography and Security.* [Preprint] https://arxiv.org/abs/2504.07839 [Accessed: 10th December 2025]

Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F. & Choo, K. (2021) Artificial intelligence in cyber security: research advances, challenges, and opportunities. *Artificial Intelligence Review.* 55, 1029–1053. https://doi.org/10.1007/s10462-021-09976-0.