Artificial Intelligence for Enhanced Cybersecurity: A Comprehensive Review and Future Directions

N. Alsakkaf [1,*]

Received: 06/08/2025
Revised: 01/09/2025
Accepted: 02/09/2025

[1] Department of Computing, faculty of Engineering and Computing, University of Science and Technology, Aden, Yemen
*Corresponding Author's Email: n.alsaqqaf@ust.edu

Artificial Intelligence for Enhanced Cybersecurity: A Comprehensive Review and Future Directions

Nasr Alsakkaf
Department of Computing, faculty of Engineering and Computing, University of Science and Technology,
Aden, Yemen
n.alsaqqaf@ust.edu

Abstract**—** The complexity and increasing rate of cyberattacks have become a formidable hurdle to people, institutions, and infrastructure vital sectors in all continents. Conventional signature-based defenses to cybersecurity threats, which are usually backreactions by nature, can hardly keep up with trends in threats that are constantly shifting. Artificial Intelligence (AI), including Machine Learning (ML) and Deep Learning (DL), is the new paradigm enabling completely new opportunities to enhance cybersecurity mechanisms. This paper is an in-depth overview of how AI can be used to enhance cybersecurity, the various uses and applications, techniques, and all other future possibilities. We also get into the various aspects of AI being used in cybersecurity, which includes intrusion detection, malware analysis, security orchestration, threat intelligence, and vulnerability management. Next, we address the challenges and limitations of the implementation of AI in cybersecurity, that is, the data privacy problems, adversarial trained AI attacks, and explainability. Lastly, we point out the emergent lines of future research with the most potential, such as blockchain + AI, further development of Explainable AI (XAI), and combining human interaction with AI. The review is intended to be a useful tool for both researchers, practitioners, and policymakers in order to learn more about AI and how it can be used to develop more proactive and resilient cybersecurity.

Keywords— cybersecurity, artificial intelligence, AI, machine learning, ML, deep learning, DL, threat detection, malware analysis, security orchestration, threat intelligence, vulnerability management, data privacy

I. INTRODUCTION

The technology age advances, and our lives become more intertwined with the digital world. While this brings incredible possibilities, it also raises tough questions and challenges that we're still figuring out. Among these, cybersecurity stands as a paramount concern, with the integrity, confidentiality, and availability of digital assets constantly under siege [1]. The global landscape of cyber threats is evolving at an alarming rate, marked by sophisticated attack vectors, polymorphic malware, advanced persistent threats (APTs), and state-sponsored cyber warfare [2]. Traditional cybersecurity paradigms, heavily reliant on predefined rules, signature-based detection, and manual human intervention, are proving increasingly inadequate in defending against these dynamic and often novel forms of cyberattacks [3]. The sheer volume of security data, coupled with the speed at which threats emerge and mutate, overwhelms human analysts and conventional security systems, leading to delayed responses and significant breaches [4].

In response to this escalating crisis, Artificial Intelligence (AI) has emerged as a revolutionary paradigm, offering a potent arsenal of tools and techniques to augment and transform cybersecurity defenses. AI, encompassing subfields such as Machine Learning (ML) and Deep Learning (DL), possesses the inherent capability to process and analyze colossal datasets, identify intricate patterns, and make intelligent decisions with a speed and scale unattainable by human cognition alone [5]. From automating routine security tasks to predicting novel threats and orchestrating complex incident responses, AI is poised to reshape the future of cybersecurity, shifting the focus from reactive defense to proactive resilience [6].

In this paper, we provide an in-depth look at the role of AI in enhancing cybersecurity. Our goal is to highlight practical applications and discuss the AI techniques driving these advancements. We also outline the challenges and the potential future directions in this rapidly evolving landscape. These are the primary objectives of this research paper:

1. To explore how AI can improve cybersecurity in areas like intrusion detection, malware analysis, and threat intelligence and vulnerability management, which will make our digital systems more secure.
2. To break down the specific AI, ML, and DL techniques being used to tackle today's cybersecurity threats, providing insight into the cutting-edge technologies seeking security advancements.
3. To critically inherent challenges and limitations of integrating AI into cybersecurity environments, examining the obstacles that come with deploying AI-driven security solutions.
4. To explore the most promising future research directions and emerging trends that will strengthen AI's role in cybersecurity, building more robust and adaptive security systems that can stay ahead of evolving threats.

The rest of this paper is organized as follows:Section II introduces the background on the cybersecurity challenges and necessary AI concepts. Section III details diverse AI technologies in cybersecurity Practical Applications of AI in Different Cybersecurity Domains (Section IV) Finally, Section V lists the AI challenges in cybersecurity. Finally, Section VI discusses future research directions and gaps. Section VII, finally, summarizes the main results and implications of the paper.

BACKGROUND

A. Evolution of Cybersecurity Challenges

Over the past few years, the cybersecurity landscape has radically changed from an early discipline in creating security for isolated systems to a worldwide priority that protects entire interconnected digital ecosystems. Initially, cybersecurity threats were relatively simplistic, primarily involving viruses and worms that exploited known vulnerabilities [7]. Defense mechanisms were largely reactive, relying on signature-based antivirus software and basic firewalls. However, with the advent of the internet, the proliferation of networked systems, and the rise of sophisticated cybercriminal organizations and nation-state actors, the nature of threats has become increasingly complex and pervasive [8].

Today's cybersecurity challenges are characterized by several key factors:

- **Sophistication of Attacks**: Modern cyberattacks, such as advanced persistent threats (APTs), polymorphic malware, fileless attacks, and highly targeted phishing campaigns, are designed to evade traditional defenses and remain undetected for extended periods [9]. These attacks often leverage zero-day vulnerabilities, making them particularly difficult to counter with signature-based methods.
- **Increased Attack Surface:** The rapid adoption of cloud computing, Internet of Things (IoT) devices, and remote work models has vastly expanded the potential attack surface, creating numerous entry points for malicious actors [10]. Each new connection has a service that represents a potential vulnerability that can be exploited.
- **Data Volume and Velocity:** The sheer volume and velocity of data generated within modern IT environments make it challenging for human analysts to monitor, analyze, and respond to security incidents effectively. Security information and event management (SIEM) systems collect massive amounts of logs and alerts, often leading to alert fatigue and missed critical events [11].
- **Evolving Threat Actors:** The motivations and capabilities of threat actors have diversified. Beyond individual hackers, organized cybercrime syndicates, state-sponsored groups, and even insider threats pose significant risks, each employing distinct tactics, techniques, and procedures (TTPs) [12].
- **Regulatory and Compliance Pressures**: Organizations face increasing pressure from regulatory bodies to protect sensitive data and report breaches, adding another layer of complexity to cybersecurity management [13].

The shifting needs are further solidified by the changing nature of threats, which explains why intelligent, adaptive, and automated cybersecurity solutions are now needed to match pace with the evolving threat landscape.

B. Fundamental Concepts of Artificial Intelligence, Machine Learning, and Deep Learning

Artificial Intelligence (AI) is a broad field of computer science dedicated to creating machines that can perform tasks typically requiring human intelligence. These tasks include learning, problem-solving, perception, and decision-making [14]. Within AI, Machine Learning (ML) and Deep Learning (DL) are especially important to cybersecurity.

Machine Learning (ML) is a subset of AI that enables systems to learn from data without being explicitly programmed. ML algorithms build a mathematical model based on sample data, known as 'training data,' in order to make predictions or decisions without being explicitly programmed to perform the task [15]. Examples for common ML tasks are 3 Classification (e.g., detecting malware), Regression (e.g., estimating risk scores), Clustering (e.g., grouping network traffic by similar characteristics), and Anomaly Detection (e.g., spotting non-natural user behavior). The primary teaching paradigms within ML are supervised, unsupervised, and reinforcement learning:

- **Supervised Learning:** Algorithms learn from labeled data, where the desired output is known. Examples include classification algorithms like Support Vector Machines (SVM), Decision Trees, Random Forests, and K-Nearest Neighbors (K-NN) [16].
- **Unsupervised Learning:** Algorithms discover patterns in unlabeled data.
- Clustering algorithms like K-Means and hierarchical clustering, and
- Dimensionality reduction techniques, like Principal Component Analysis (PCA), are common examples [17].
- **Reinforcement Learning (RL):** Agents learn to make decisions by performing actions in an environment to maximize a cumulative reward. RL is particularly useful for dynamic and adaptive security tasks [18].
- **Deep Learning (DL)** is a specialized subfield of ML that uses artificial neural networks with multiple layers (hence 'deep') to learn representations of data with multiple levels of abstraction [19]. DL models are capable of learning complex patterns directly from raw data, eliminating the need for manual feature engineering, which is a significant advantage in domains like cybersecurity where relevant features can be obscure or high-dimensional [20]. Key DL architectures include:
- **Convolutional Neural Networks (CNNs)**: Primarily used for image and video analysis, but also applicable to network traffic analysis by treating data as images [21].
- **Recurrent Neural Networks (RNNs):** Designed for sequential data, such as network packet sequences or log files, including variants like Long Short-Term Memory (LSTM) networks [22].
- **Generative Adversarial Networks (GANs):** Consist of a generator and a discriminator network that compete against each other, useful for generating synthetic data for training or detecting sophisticated malware [23].

C. Overview of the NIST Cybersecurity Framework

A voluntary framework for managing cybersecurity risks, the National Institute of Standards and Technology's (NIST) Cybersecurity Framework (CSF) consists of best practices, guidelines, and standards. It provides a common language and systematic approach for organizations to assess and improve their cybersecurity posture [24]. The framework is structured around five core functions, which provide a high-level strategic view of an organization's management of cybersecurity risk:

1. **Identify**: Develop an organizational understanding to manage cybersecurity risk to systems, assets, data, and capabilities. This function involves activities such as asset management, business environment understanding, governance, risk assessment, and risk management strategy [25].

2. **Protect**: Develop and implement appropriate safeguards to ensure delivery of critical infrastructure services. This includes access control, awareness and training, data security, information protection processes and procedures, maintenance, and protective technology [26].
3. **Detect:** Develop and implement appropriate activities to identify the occurrence of a cybersecurity event. This involves anomalies and events, continuous security monitoring, and detection processes [27].
4. **Respond:** Develop and implement appropriate activities to take action regarding a detected cybersecurity incident. This covers response planning, communications, analysis, mitigation, and improvements [28].
5. **Recover:** Develop and implement appropriate activities to maintain plans for resilience and to restore any capabilities or services that were impaired due to a cybersecurity incident. This includes recovery planning, improvements, and communications [29].

These five functions can be used to map AI applications in cybersecurity, improving capabilities at every phase of the cybersecurity lifecycle. For instance, AI can assist in identifying vulnerabilities, protecting systems through intelligent access controls, detecting anomalies in real-time, automating incident response, and aiding in recovery efforts by analyzing post-incident data [30].

II. AI Techniques for Enhanced Cybersecurity

Artificial intelligence has the potential to revolutionize cybersecurity because it can use a wide range of sophisticated computational methods. These methods greatly enhance conventional security measures by empowering systems to learn from enormous datasets, spot intricate patterns, and make wise decisions. The main AI methods that are transforming cybersecurity are covered in detail in this section.

D. Machine Learning Algorithms

The foundation of many AI-driven cybersecurity solutions is machine learning (ML). By learning from past data, its algorithms are skilled at spotting irregularities, categorizing threats, and forecasting malevolent activity. Important machine learning algorithms commonly used in cybersecurity include

- **Support Vector Machines (SVM**): SVMs are powerful supervised learning models used for classification and regression analysis. In cybersecurity, SVMs are effective in distinguishing between legitimate and malicious network traffic, classifying malware families, and detecting intrusions by finding an optimal hyperplane that separates different classes of data points [31]. Their capability to handle high-dimensional data makes them suitable for complex security datasets.
- **Decision Trees and Random Forests:** Choosing Trees are structures that resemble flowcharts, with each internal node standing for an attribute test, each branch for the test's result, and each leaf node for a class label. An ensemble learning technique called Random Forests builds several decision trees during training and produces a class that is the mean prediction (regression) or the mode of the classes (classification) of the individual trees. These algorithms are widely used for intrusion detection, spam filtering, and fraud detection due to their interpretability and ability to handle both numerical and categorical data [32].
- **K-Nearest Neighbors (K-NN**): K-NN is a non-parametric, instance-based learning algorithm used for classification and regression. In cybersecurity, K-NN can be applied for anomaly detection by identifying data points that are significantly different from their k-nearest neighbors, which can indicate a potential cyberattack or unusual system behavior [33].
- **Clustering Algorithms (e.g., K-Means, DBSCAN):** Finding organic clusters or groupings in unlabeled security data is made possible by unsupervised learning methods like K-Means and DBSCAN. This is particularly useful for discovering new types of attacks, segmenting network traffic, or grouping similar malicious activities without prior knowledge of their characteristics [34]. K-Means, for example, can group network connections according to their characteristics, enabling security analysts to spot unusual groupings that could indicate a cyberattack.

E. Deep Learning Architectures

Deep Learning (DL), a branch of machine learning (ML), is particularly effective at tasks involving intricate patterns and vast amounts of unprocessed data because it uses multi-layered neural networks to learn hierarchical representations of data. DL architectures are particularly effective in cybersecurity, where traditional ML methods might struggle with feature engineering or high-dimensional data [35].

- **Convolutional Neural Networks (CNNs): CNNs** have found important uses in cybersecurity, despite their traditional use in image processing. They can analyze network traffic data by converting it into a 2D image-like format, enabling the detection of sophisticated network intrusions and malware patterns [36]. By treating byte sequences as images, CNNs are also used to analyze binary code for malware classification.
- **Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM):** Because RNNs are made to handle sequential data, they are perfect for examining time-series data, including system calls, network logs, and user behavior patterns. LSTMs, a special type of RNN, address the vanishing gradient problem and are highly effective in capturing long-term dependencies in sequential data, crucial for detecting advanced persistent threats and anomalous user activities over time [37].
- **Generative Adversarial Networks (GANs):** GANs are made up of two neural networks that compete with one another: a discriminator and a generator. In cybersecurity, GANs can be used to generate synthetic malicious samples for training more robust detection models or, conversely, to detect sophisticated polymorphic malware that constantly changes its signature to evade detection [38]. They can also be used to test the resilience of security systems by simulating adversarial attacks.

- **Autoencoders:** Unsupervised neural networks called autoencoders are employed in anomaly detection and dimensionality reduction. By learning to reconstruct their input, autoencoders can identify anomalies as data points that have high reconstruction errors, indicating they deviate significantly from normal patterns. This makes them valuable for detecting novel attacks or unusual system states [39].

1. Federated Learning (FL)

- Federated Learning (FL) is a decentralized ML approach that enables multiple entities (e.g., organizations, devices) to collaboratively train a shared model without exchanging their raw data [40]. Instead, data privacy and confidentiality are maintained by sharing only model updates (such as weight changes). FL is especially helpful in cybersecurity for:
- **Collaborative Threat Detection:** Using their local datasets, several security organizations can work together to train a more reliable threat detection model without revealing patient data or confidential proprietary information. This allows for the detection of widespread threats more effectively [41].
- **Privacy-Preserving Malware Classification:** FL can facilitate the development of highly accurate malware classification models by leveraging diverse datasets from various sources, while ensuring that individual malware samples or sensitive system information remain localized [42].
- **IoT Security**: With the proliferation of IoT devices, FL offers a scalable and privacy-preserving method to train security models on edge devices, enabling real-time anomaly detection and threat intelligence without centralizing vast amounts of sensitive IoT data [43].

2. Explainable AI (XAI)

- As AI models become more complex, particularly deep learning models, their decision-making processes can become opaque, leading to what is often referred to as the "black box" problem. Explainable AI (XAI) aims to make AI models more transparent, interpretable, and understandable to humans [44]. In cybersecurity, XAI is critical for:
- **Building Trust and Confidence:** To trust an AI model's recommendations and take the necessary action, security analysts must comprehend why the model identified a specific activity as malicious. XAI provides insights into the model's reasoning, fostering confidence in AI-driven security solutions [45].
- **Improving Incident Response:** Effective incident response and forensic analysis depend on knowing the underlying cause of a security incident and the elements that contributed to the AI's detection. XAI can help pinpoint critical features or events that contributed to the alert [46].
- **Regulatory Compliance:** Organizations may be asked to prove the equity and openness of their AI systems as laws pertaining to data privacy and AI ethics become stricter. XAI can assist in meeting these compliance requirements [47].
- **Debugging and Model Improvement**: By understanding why an AI model makes certain predictions, security researchers can identify biases, errors, or limitations in the model and iteratively improve its performance and robustness [48].
- **Techniques for XAI include:**
- o Feature Importance: determining which input features have the biggest influence on a model's forecast.
- o Local Interpretable Model-agnostic Explanations (LIME): Using an interpretable model to locally approximate the predictions of any classifier or regressor.
- o Shapley Additive explanations (SHAP): an explanation of any machine learning model's output using game theory.

3. Reinforcement Learning and Other Advanced AI Paradigms

- **Reinforcement Learning (RL)** entails teaching agents how to behave in a given environment in order to maximize a cumulative reward. RL can be used in cybersecurity for:
- **Automated Penetration Testing**: RL agents can learn to navigate complex network environments, identify vulnerabilities, and exploit them, mimicking the behavior of real attackers to test system resilience [49].
- **Adaptive Defense Systems**: RL can enable security systems to adapt their defense strategies in real-time based on observed attack patterns and system states, leading to more dynamic and proactive protection [50].
- **Malware Obfuscation and De-obfuscation**: RL can be used to develop techniques for malware to evade detection or, conversely, for security systems to de-obfuscate complex malware [51].
- Other innovative AI paradigms gaining traction in cybersecurity include:
- **Natural Language Processing (NLP)**: For analyzing unstructured text data such as security reports, threat intelligence feeds, and social media to identify emerging threats, phishing attempts, and sentiment analysis related to cybersecurity incidents [52].
- **Computer Vision**: For analyzing visual data, such as screenshots of malicious websites, phishing emails, or even physical security footage, to detect anomalies or identify threats [53].
- **Graph Neural Networks (GNNs)**: For analyzing complex relationships in network data, such as user-device connections, communication patterns, and attack graphs, to identify sophisticated threats that involve multiple entities [54].

III. Applications of AI in Cybersecurity

- Artificial intelligence is revolutionizing cybersecurity in a number of ways, providing creative answers to persistent issues and opening up new defensive capabilities. The main uses of AI in various cybersecurity domains are examined in this section.

4. Intrusion Detection Systems (IDS)

- Any cybersecurity infrastructure must include intrusion detection systems (IDS), which are made to keep an eye on system or network activity for malicious activity or policy infractions. Conventional intrusion detection systems frequently use signature-based detection, which works well against known threats but is unable to detect

new or zero-day attacks. AI-powered IDSs, particularly those leveraging ML and DL, offer a more adaptive and proactive approach [55].

- **Anomaly Detection**: A network, system, or user's typical behavior patterns can be learned by AI models. An anomaly is identified as any notable departure from these learned baselines, which may be a sign of malicious activity or an intrusion. Unsupervised learning algorithms like clustering (e.g., K-Means, DBSCAN) and autoencoders are frequently used for this purpose, as they do not require pre-labeled attack data [56]. Supervised learning methods, such as SVMs and Random Forests, can also be trained on labeled datasets of normal and anomalous traffic to classify new events [57].
- **Signature-based Detection Enhancement**: While AI excels at anomaly detection, it can also enhance traditional signature-based methods. ML can automatically generate and update signatures for new malware variants or attack patterns by analyzing large volumes of threat data, reducing the manual effort required for signature creation [58].

5. **Network Traffic Analysis**: DL models, particularly CNNs and RNNs, are highly effective in analyzing raw network packet data or flow records to identify subtle patterns indicative of attacks like Distributed Denial of Service (DDoS), port scanning, or sophisticated multi-stage intrusions [59]. DL can reveal hidden correlations that human analysts might overlook by treating network traffic as sequential data or even transforming it into representations that resemble images.

6. Malware Analysis and Detection
- Malware, which includes ransomware, worms, viruses, and rootkits, is still a constant and changing threat. AI plays a crucial role in enhancing malware analysis and detection capabilities, moving beyond static signature matching to more dynamic and behavioral analysis [60].
- Static Analysis: Without running the code, ML algorithms can identify executable files as malicious or benign by examining their static characteristics, such as header data, string patterns, and API calls. This approach is fast and can identify known malware families [61].
- Dynamic Analysis (Behavioral Analysis): In a sandbox setting, AI models can track suspicious files' runtime behavior, noting how they interact with the operating system, make network connections, and alter the file system. DL models, especially RNNs, are adept at identifying malicious behavioral sequences, even for polymorphic or obfuscated malware that constantly changes its code to evade detection [62].
- Zero-Day Threat Detection: By focusing on anomalous behavior rather than known signatures, AI-driven malware detection systems are better equipped to identify previously unseen (zero-day) malware, which poses a significant challenge to traditional antivirus solutions [63].
- Malware Family Classification: AI can automatically group new malware samples into known families based on their characteristics and behavior, aiding in threat intelligence and enabling more targeted defense strategies [64].

7. Security Orchestration, Automation, and Response (SOAR)
- Security teams can react to incidents more quickly thanks to Security Orchestration, Automation, and Response (SOAR) platforms, which automate repetitive tasks and integrate multiple security tools. AI significantly enhances SOAR capabilities by providing intelligence and decision-making capabilities [65].
- **Automated Incident Response**: AI is able to correlate events, analyze security alerts from multiple sources (such as firewalls, SIEMs, and IDS), and automatically initiate pre-established response actions, like blocking malicious IP addresses, isolating compromised hosts, or starting forensic data collection. This reduces response times from hours to minutes or even seconds [66].
- **Threat Prioritization**: Security analysts can concentrate on the most important threats first because AI can rank incidents according to their likelihood, potential impact, and severity given the deluge of security alerts. ML models can learn from historical incident data to assign accurate risk scores [67].
- **Playbook Optimization**: Based on new threat trends and effective previous responses, AI can evaluate the efficacy of various response playbooks and recommend improvements or create new playbooks. Reinforcement learning can be particularly useful in this context, as it can learn optimal response strategies through trial and error [68].
- **Natural Language Processing (NLP) for Incident Triage**: NLP can be used to process unstructured data from security tickets, emails, and threat intelligence feeds, extracting key information to automate incident triage and routing [69].

8. Threat Intelligence and Prediction
- Information about potential and real threats to an organization is gathered, processed, and analyzed as part of threat intelligence. AI significantly enhances threat intelligence by automating data collection, identifying emerging trends, and predicting future attacks [70].
- **Automated Data Collection and Analysis**: AI can gather data on new vulnerabilities, attack campaigns, and threat actor TTPs by continuously monitoring enormous volumes of open-source intelligence (OSINT), dark web forums, social media, and proprietary threat feeds. NLP is crucial for extracting meaningful insights from unstructured text data [71].
- **Predictive Analytics**: To forecast the possibility of future attacks, find possible targets, and foresee the kinds of threats an organization may encounter, machine learning models can examine historical attack data, vulnerability reports, and geopolitical events. This enables proactive defense strategies and resource allocation [72].
- **Threat Actor Profiling**: AI can build profiles of threat actors by analyzing their past activities, tools, and infrastructure, helping security teams understand their adversaries and anticipate their next moves [73].
- Vulnerability Prioritization: By correlating threat intelligence with an organization's asset inventory and

vulnerability scan results, AI can prioritize vulnerabilities based on their exploitability and potential impact, guiding patch management efforts [74].

1. Vulnerability Management
- The ongoing process of locating, evaluating, ranking, and fixing security flaws in systems and applications is known as vulnerability management. AI can significantly streamline and enhance this process [75].
- **Automated Vulnerability Scanning and Analysis:** It is the vulnerability management that continues the process of identifying, assessing, prioritizing, and remediating security weaknesses in the systems and applications. ML can analyze scan results and prioritize vulnerabilities based on real-world exploitability and business impact [76].
- **Predictive Vulnerability Scoring**: Instead of relying solely on static Common Vulnerability Scoring System (CVSS) scores, AI can develop dynamic risk scores for vulnerabilities by incorporating real-time threat intelligence, asset criticality, and exploit availability, providing a more accurate picture of risk [77].
- **Patch Management Optimization**: AI can analyze dependencies, system uptime requirements, and potential conflicts to recommend optimal patch deployment schedules, minimizing disruption while maximizing security [78].

2. User and Entity Behavior Analytics (UEBA)
- User and Entity Behavior Analytics (UEBA) focuses on detecting insider threats, targeted attacks, and financial fraud by analyzing the behavior of users and other entities (e.g., applications, devices) within an organization. AI is fundamental to UEBA solutions [79].
- **Baseline Behavior Profiling**: By examining a variety of data sources, such as login patterns, access times, data transfer volumes, and application usage, AI models create baselines of typical behavior for every user and entity.
- Unsupervised learning is often employed here to discover normal patterns without predefined rules [80].
- **Anomaly Detection**: An alert is triggered by any notable departure from the defined baseline behavior, such as odd login locations, access to private information outside of business hours, or excessive data downloads. AI can differentiate
- between legitimate deviations and malicious activities, reducing false positives [81].
- **Insider Threat Detection**: AI-powered UEBA is especially good at spotting insider threats, which occur when authorized users abuse their access rights for nefarious ends. By detecting subtle changes in behavior, AI can flag suspicious activities that might otherwise go unnoticed [82].

3. Network Security
- AI enhances various aspects of network security beyond just intrusion detection, contributing to more robust and adaptive network defenses [83].
- **Traffic Classification and Filtering**: Even when conventional port-based or signature-based techniques are unsuccessful, AI can reliably identify network traffic (such as malicious traffic, encrypted tunnels, and legitimate applications). This enables more granular filtering and policy enforcement [84].
- **DDoS Attack Mitigation**: AI can detect the subtle precursors of DDoS attacks and differentiate between legitimate traffic surges and malicious floods, enabling faster and more effective mitigation strategies [85].
- **Secure Network Orchestration**: AI can dynamically reconfigure network policies, segment networks, and adjust security controls in response to detected threats or changes in network conditions, creating a more resilient and self-healing network [86].

IV. Challenges and Limitations of AI in Cybersecurity

Artificial intelligence has the potential to completely transform cybersecurity, but there are a number of obstacles and restrictions that must be overcome before it can be put into practice. For AI to be successfully and responsibly implemented in vital security infrastructures, these problems must be resolved.

A. Data Privacy and Ethical Concerns

AI models are data-hungry, especially machine learning-based models. Access to enormous volumes of sensitive data, such as network traffic logs, user behavior data, and incident reports, is frequently necessary for training efficient cybersecurity models. This raises significant privacy concerns, especially when dealing with personally identifiable information (PII) or confidential business data [87].
- **Data Collection and Storage**: Strict data protection laws (such as the CCPA and GDPR) must be followed when gathering, storing, and processing vast amounts of sensitive data for AI training. Ensuring data anonymization, pseudonymization, and secure storage mechanisms is paramount [88].
- **Bias in Data**: The AI model will pick up and reinforce biases from the training data, which could result in unfair treatment or discriminatory results. For instance, an AI system trained on data primarily from one demographic might misidentify legitimate activities from another demographic as suspicious [89].
- **Ethical Implications of Autonomous Decisions**: As AI systems gain more autonomy in making security decisions (e.g., blocking access, isolating systems), ethical questions arise regarding accountability, transparency, and the potential for unintended consequences. Who is responsible when an autonomous AI system makes a wrong decision that leads to a security breach or denies legitimate access? [90]

4. Adversarial AI Attacks

One of the most significant challenges for AI in cybersecurity comes from adversarial AI attacks, where malicious actors intentionally manipulate input data to deceive AI models [91]. AI-driven security systems may become less effective as a result of these attacks.
- Evasion Attacks: Attackers create malicious inputs (such as malware samples and network packets) that are intended to be mistakenly identified by the AI model as benign in order to avoid detection. This is often achieved

by adding small, imperceptible perturbations to the input data [92].
- Poisoning Attacks: Attackers introduce malicious data into an AI model's training dataset, tainting the learning process and leading to inaccurate predictions or particular malicious behaviors when the model is deployed. This can be particularly damaging if the attacker gains control over the training data pipeline [93].
- Model Inversion Attacks: Attackers attempt to reconstruct sensitive training data from the deployed AI model, potentially compromising privacy [94].
- Model Extraction Attacks: Attackers try to steal the underlying AI model or its parameters, which can then be used to launch more effective adversarial attacks or to replicate proprietary AI capabilities [95].

5. Explainability and Interpretability Issues (Black-Box Models)
- Many powerful AI models, particularly deep learning networks, function as "black boxes," which means that their internal decision-making processes are opaque and challenging for humans to comprehend, as was covered in Section III. This lack of transparency poses significant challenges in cybersecurity [96].
- **Lack of Trust**: If security analysts and incident responders are unable to comprehend the reasoning behind a specific alert or suggested course of action, they may be reluctant to fully trust or rely on AI systems. As a result, adoption may decline, and human intuition may become less important than AI insights.
- **Difficulty in Debugging and Auditing**: In a black-box model, it can be difficult to identify the precise cause of an AI model's error when it predicts something incorrectly or fails to recognize a threat. This makes debugging, auditing, and improving the model a complex task [97].
- **Compliance and Accountability**: Organizations may be required to prove the impartiality, dependability, and reasoning behind their security systems in regulated sectors. Black-box AI models can hinder compliance efforts and make it difficult to assign accountability in the event of a security failure [98].

6. Resource Intensity and Computational Overhead
- Powerful GPUs, large memory capacities, and significant energy consumption are just a few of the computational resources needed to train and implement complex AI models, particularly deep learning architectures. This can be a barrier for smaller organizations or those with limited IT infrastructure [99].
- **Training Time**: Rapid iteration and deployment are difficult because it can take days or even weeks to train complex DL models on large cybersecurity datasets.
- **Inference Latency**: Real-time cybersecurity applications require very low latency, even though inference (making predictions) is typically faster than training.
- Complex models might introduce delays that are unacceptable for critical security operations [100].
- **Cost**: Some organizations may not be able to afford the costly hardware and software infrastructure needed for AI development and implementation.

7. Data Imbalance and Quality Issues
- Cybersecurity datasets often suffer from inherent challenges that can negatively impact the performance of AI models [101].

Data **Imbalance**: Malicious activities are typically rare compared to legitimate activities, leading to highly imbalanced datasets. Training AI models on such datasets can result in models that are biased towards the majority class (normal behavior) and perform poorly in detecting the minority class (attacks) [102]. To deal with this, methods like oversampling, undersampling, or creating synthetic data are frequently needed.
Lack **of Labeled Data**: One of the biggest obstacles is getting high-quality, labeled cybersecurity data. Classifying malware, or determining whether network traffic is malicious or benign, is a costly and time-consuming procedure that calls for specialized knowledge. This scarcity of labeled data can limit the effectiveness of supervised learning approaches [103].
Data **Quality and Noise**: Cybersecurity data may contain errors or irrelevant features and be noisy, inconsistent, and incomplete. Poor data quality can lead to inaccurate models and unreliable predictions [104].
Concept **Drift**: Because cyber threats are always changing, the patterns that an AI model learns may eventually become out of date. This phenomenon, known as concept drift, requires continuous retraining and adaptation of AI models to maintain their effectiveness [105].

Future Directions and Research Gaps
Because AI and cybersecurity are developing so quickly, it is imperative that ongoing research and innovation be done to solve new problems and fully utilize AI's potential for protecting digital assets. This section highlights important research gaps that require more study and suggests promising future directions.
Integration of Blockchain with AI for Enhanced Security and Data Integrity Blockchain technology, with its inherent properties of decentralization, immutability, and transparency, offers a compelling solution for enhancing the security and integrity of AI systems in cybersecurity [106].
Secure Data Sharing for Federated Learning: Blockchain can provide a secure and auditable mechanism for sharing model updates in federated learning environments, ensuring the integrity of the training process and preventing malicious tampering [107].
Decentralized Threat Intelligence: A blockchain-based platform could enable secure and verifiable sharing of threat intelligence among organizations, fostering collaborative defense against cyberattacks without relying on a centralized authority [108].
Tamper-Proof AI Models: Storing AI model parameters or hashes on a blockchain could ensure the integrity and authenticity of deployed models, protecting against model poisoning or unauthorized modifications [109].
Secure IoT Ecosystems: The combination of AI for anomaly detection and blockchain for secure device identity and communication can create more resilient and trustworthy security solutions [110].
Research Gap: It's still very difficult to create scalable and effective blockchain-AI integration frameworks that can manage the massive transaction volumes and processing

demands of real-time cybersecurity operations. Consensus procedures and smart contract designs tailored for these hybrid systems require more investigation.

B. Advanced XAI Techniques for Critical Security Decisions

While current XAI techniques provide valuable insights, there is a need for more sophisticated and context-aware explainability methods, especially for critical security decisions where human trust and understanding are paramount [111].

Actionable Explanations: Future XAI research should focus on generating explanations that are not only understandable but also actionable, guiding security analysts on what steps to take based on the AI's reasoning [112].

Counterfactual Explanations: Providing explanations that show what would have to change in the input for the AI to make a different decision can be highly valuable for understanding vulnerabilities and attack vectors [113].

Human-in-the-Loop XAI: Developing interactive XAI systems that allow security experts to query the AI model, refine explanations, and provide feedback to improve model performance and interpretability [114].

Research Gap: It is essential to bridge the gap between the technical justifications produced by XAI tools and the real-world knowledge needed by various security professionals (such as incident responders, forensic analysts, and compliance officers). Methodologies for user-centric XAI design and assessment require further study.

C. Development of Robust AI Models Against Adversarial Attacks

The vulnerability of AI models to adversarial attacks poses a serious threat to their deployment in cybersecurity. Future research must focus on developing more robust and resilient AI models [115].

Adversarial Training: Incorporating adversarial examples into the training data to make models more robust to evasion attacks [116].

Defensive Distillation: Training a second model on the softened outputs of an initial model to reduce the susceptibility to adversarial perturbations [117].

Certified Robustness: Developing methods to mathematically guarantee the robustness of AI models against certain types of adversarial attacks [118].

Proactive Adversarial Defense: Research into techniques that can detect and mitigate adversarial attacks in real-time before they impact the AI model's performance [119].

Research Gap: Even though there has been progress, it is still difficult to achieve certified robustness against a variety of adversarial attacks in intricate, high-dimensional cybersecurity datasets. Future research should focus on creating adversarial mechanisms that are both realistic and computationally effective.

D. AI in Quantum-Safe Cryptography

As quantum computing advances, current cryptographic standards are at risk. AI can play a role in the transition to quantum-safe cryptography [120].

Post-Quantum Cryptography (PQC) Analysis: AI can assist in analyzing the security and efficiency of new PQC algorithms against classical and quantum attacks [121].

Quantum Random Number Generation: AI can be used to enhance the quality and randomness of quantum random number generators, which are crucial for strong cryptographic keys [122].

AI for Quantum Key Distribution (QKD) Optimization: Optimizing QKD networks and protocols using AI to improve their efficiency and security [123].

Research Gap: Cybersecurity at the nexus of AI and quantum computing is a young field. To find new vulnerabilities that could result from this convergence and to comprehend how AI can effectively aid in the creation and implementation of quantum-safe cryptographic solutions, more research is required.

E. Human-AI Collaboration in Cybersecurity Operations

Instead of replacing human analysts, AI should be seen as a powerful augmentation tool. Future research should focus on optimizing human-AI collaboration to leverage the strengths of both [124].

Intelligent Assistants: Developing AI-powered assistants that can provide real-time insights, automate mundane tasks, and guide human analysts through complex investigations [125].

Shared Understanding and Mental Models: Research into how humans and AI can develop shared mental models of the cybersecurity landscape and ongoing threats to facilitate more effective collaboration [126].

Adaptive Learning Systems: AI systems that can learn from human feedback and adapt their behavior to better support human decision-making [127].

Research Gap: One crucial area of research is creating efficient human-AI interfaces and interaction paradigms that promote smooth cooperation and shield human analysts from cognitive overload. It's also critical to comprehend the sociological and psychological facets of human-AI collaboration in high-stakes cybersecurity settings.

Conclusion

While the digital world presents unmatched opportunities, it is also rife with an increasing number of sophisticated and persistent cyberthreats. In the face of dynamic and polymorphic attacks, the shortcomings of conventional, signature-based cybersecurity guards have highlighted the pressing need for more proactive, intelligent, and adaptive solutions. With its revolutionary potential to improve all aspects of cybersecurity, artificial intelligence—which includes machine learning and deep learning—has become a key technology.

This thorough analysis has brought to light AI's significant influence in a number of cybersecurity domains. We have looked at how anomaly detection and sophisticated network traffic analysis are used by AI-driven intrusion detection systems (IDS) to find new threats. AI in malware analysis goes beyond static signatures to dynamic behavioral analysis, which makes it possible to identify advanced polymorphic malware and zero-day threats. AI greatly enhances Security

Orchestration, Automation, and Response (SOAR) platforms, resulting in intelligent threat prioritization, optimized playbooks, and automated incident response. Additionally, by automating data collection, enabling predictive analytics, and profiling threat actors, AI transforms threat intelligence. AI supports automated scanning, predictive scoring, and patch management optimization in vulnerability management. By creating and tracking behavioral baselines, AI-powered User and Entity Behavior Analytics (UEBA) offers vital capabilities for identifying targeted attacks and insider threats. Lastly, through secure network orchestration, intelligent traffic classification, and DDoS mitigation, AI helps to ensure strong network security.

Despite these impressive developments, there are still difficulties in integrating AI into cybersecurity. Black-box models' intrinsic clarity of the problems, vulnerability to adversarial AI attacks (evasion and poisoning), and worries about data privacy and ethical ramifications continue to be significant obstacles. The widespread adoption and optimal performance of AI in real-world cybersecurity environments are further complicated by resource intensity, computational overhead, and issues with data imbalance and quality.

The future of AI in cybersecurity looks good, but more research and development is needed. Some promising directions are to combine blockchain technology with AI in a way that protects data integrity and makes it safe to work together. Learning; the advancement of sophisticated Explainable AI (XAI) methodologies that yield actionable and comprehensible insights for security experts; and the formulation of AI models that are intrinsically robust and resilient to complex adversarial assaults. The new field of AI in quantum-safe cryptography has a lot of potential for making our digital guards more secure in the future. Also, making sure that AI and people can work together well will be very important for using AI as a tool to help people instead of replacing their skills.

In conclusion, artificial intelligence is not merely an incremental improvement but a foundational shift in how we approach cybersecurity. By harnessing its power, we can move towards a more intelligent, automated, and adaptive defense posture, capable of anticipating and neutralizing the increasingly complex threats of the digital age. The journey is ongoing, and continuous innovation at the intersection of AI and cybersecurity will be vital in building a more secure and resilient digital future.

REFERENCES

[1] S. M. R. Islam, M. S. Rahman, M. M. Islam, and M. A. Hossain, "Cybersecurity challenges and solutions in the era of Industry 4.0: A review," J. Netw. Comput. Appl., vol. 196, p. 103234, Dec. 2021.

[2] E. Al-Shaer and S. Al-Haj, "A survey on cyber-physical system security: Threats, attacks, and solutions," Comput. Secur., vol. 100, p. 102071, Jan. 2021.

[3] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," J. Netw. Comput. Appl., vol. 196, p. 103234, Dec. 2021.

[4] R. Kaur, D. GabrijelcVicV, and T. KlobucVar, "Artificial intelligence for cybersecurity: Literature review and future research directions," Inf. Fusion, vol. 97, p. 101804, Sep. 2023.

[5] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.

[6] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.

[7] A. Aldweesh, M. Al-Rodhaan, and A. Al-Dhelaan, "A survey on cybersecurity threats and solutions in cloud computing," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 1, pp. 1-19, Jan. 2021. [8] S. M. R. Islam, M. S. Rahman, M. M. Islam, and M. A. Hossain, "Cybersecurity challenges and solutions in the era of Industry 4.0: A review,"
J. Netw. Comput. Appl., vol. 196, p. 103234, Dec. 2021. [9] A. Al-Garadi, M. A. Mohamed,
A. K. Al-Obaidi, and A. K. Al-Dhelaan, "A survey on cyber-physical system security: Threats, attacks, and solutions," *Comput. Secur.*, vol. 100, p. 102071, Jan. 2021.

[10] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021. [11] R. Kaur, D. GabrijelcVicV, and T. KlobucVar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.

[12] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.

[13] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015.

[15] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, Sep. 1995.

[17] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, vol. 1, 1967, pp. 281-297.

[18] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436- 444, May 2015.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[21] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097-1105.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[23] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672-2680.

[24] National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1, NIST, Gaithersburg, MD, USA, Apr. 2018. [25] National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1, NIST, Gaithersburg, MD, USA, Apr. 2018. [26] National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1, NIST, Gaithersburg, MD, USA, Apr. 2018. [27] National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1, NIST, Gaithersburg, MD, USA, Apr. 2018. [28] National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1, NIST, Gaithersburg, MD, USA, Apr. 2018. [29] National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1, NIST, Gaithersburg, MD, USA, Apr. 2018.
[30] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[31] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, Sep. 1995. [32] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, Oct. 2001. [33] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1,
pp. 21-27, Jan. 1967.
[34] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, vol. 1, 1967, pp. 281-297.
[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015. [36] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097-1105.
[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
[38] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672-2680.
[39] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096-1103.
[40] J. Kairouz *et al.*, "Federated learning: Challenges, methods, and future directions," *Found. Trends Mach. Learn.*, vol. 13, no. 2, pp. 89-198, 2021.
[41] Q. Yang *et al.*, "Federated learning for mobile edge intelligence: A survey," *Proc. IEEE*, vol. 108, no. 8, pp. 1359-1392, Aug. 2020.
[42] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024. [43] S. Lim, J. Kim, and J. Lee, "Federated learning-driven cybersecurity framework for IoT edge devices," *IEEE Access*, vol. 9, pp. 10203-10214, 2021.
[44] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2nd ed. 2022.
[45] F. D. S. de Oliveira, A. S. de Oliveira, and L. P. de Lima, "Explainable AI for cybersecurity: A survey," *Comput. Secur.*, vol. 106, p. 102299, Jul. 2021.
[46] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[47] F. D. S. de Oliveira, A. S. de Oliveira, and L. P. de Lima, "Explainable AI for cybersecurity: A survey," *Comput. Secur.*, vol. 106, p. 102299, Jul. 2021.
[48] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2nd ed. 2022.
[49] H. Al-Ameri, M. Al-Rodhaan, and A. Al-Dhelaan, "Reinforcement learning for automated penetration testing: A survey," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[50] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[51] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p.101804, Sep. 2023.
[52] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques,"
J. Big Data, vol. 11, no. 1, p. 105, Aug. 2024.
[53] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review,"*J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[54] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015.
[55] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[56] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, vol. 1, 1967, pp. 281-297.
[57] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, Sep. 1995.
[58] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, Oct. 2001.
[59] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097-1105.
[60] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI- driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[61] R. Kaur,D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.

[62] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
[63] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672-2680.
[64] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096- 1103.
[65] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[66] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[67] R. Kaur, D. GabrijelcVc, and T. KlobucaYr, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[68] H. Al-Ameri, M. Al- Rodhaan, and A. Al-Dhelaan, "Reinforcement learning for automated penetration testing: A survey," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[69] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[70] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p.102693, Apr. 2021.
[71] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A
comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[72] R. Kaur, D. GabrijelcVc, and T. KlobucaYr, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[73] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[74] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[75] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[76] R. Kaur, D. GabrijelcVc, and T. KlobucaYr, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[77] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[78] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[79] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[80] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, vol. 1, 1967, pp. 281-297.
[81] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, Sep. 1995.
[82] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, Oct. 2001.
[83] R. Kaur, D. GabrijelcVc, and T. KlobucaYr, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[84] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024. [85] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021. [86] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A.
Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021. [87]
R. Kaur, D. GabrijelcVc, and T. KlobucaYr, "Artificial intelligence for cybersecurity:
Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023. [88] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[89] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[90] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[91] N. Papernot, P. McDaniel, A. Sinha, and M. Fredrikson, "Towards the science of security for machine learning," *arXiv preprint arXiv:1611.03814*, 2016.
[92] C. Szegedy *et al.*, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.
[93] B. Biggio, B. Nelson, and F. Roli, "Poisoning attacks against support vector machines," in *Proc. 28th Int. Conf. Mach. Learn.*, 2012, pp. 1807-1814.
[94] M. Fredrikson, S. Jha, and T. Ristenpart, "Privacy in machine learning: Analyzing and protecting reconstruction and attribute inference attacks," in *Proc. 2015 IEEE Symp. Secur. Privacy (SP)*, 2015, pp. 575-592.
[95] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction APIs," in *Proc. 25th USENIX Secur. Symp.*, 2016, pp. 601-618. [96] C. Molnar, *Interpretable Machine*

Learning: A Guide for Making Black Box Models Explainable. 2nd ed. 2022.
[97] F. D. S. de Oliveira, A. S. de Oliveira, and L. P. de Lima, "Explainable AI for cybersecurity: A survey," *Comput. Secur.*, vol. 106,
p. 102299, Jul. 2021.
[98] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques,"
J. Big Data, vol. 11, no. 1, p. 105, Aug. 2024.
[99] R. Kaur, D. Gabrijelcv, and T. Klobucar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[100] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[101] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[102] R. Kaur, D. Gabrijelcv, and T. Klobucar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[103] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques,"
J. Big Data, vol. 11, no. 1, p. 105, Aug. 2024.
[104] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[105] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.
[106] Y. Lu and X. Li, "Blockchain and AI for cybersecurity: A survey," *IEEE Access*, vol. 8, pp. 138380-138392, 2020.
[107] Q. Yang *et al.*, "Federated learning for mobile edge intelligence: A survey," *Proc. IEEE*, vol. 108, no. 8, pp. 1359-1392, Aug. 2020. [108] Y. Lu and X. Li, "Blockchain and AI for cybersecurity: A survey," *IEEE Access*, vol. 8, pp. 138380-138392, 2020.
[109] S. Lim, J. Kim, and J. Lee, "Federated learning-driven cybersecurity framework for IoT edge devices," *IEEE Access*, vol. 9, pp. 10203-10214, 2021.
[110] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[111] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2nd ed. 2022.
[112] F. D. S. de Oliveira,
A. S. de Oliveira, and L. P. de Lima, "Explainable AI for cybersecurity: A survey," *Comput. Secur.*, vol. 106, p. 102299, Jul. 2021.
[113] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI- driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[114] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2nd ed. 2022.
[115] N. Papernot, P. McDaniel, A. Sinha, and M. Fredrikson, "Towards the science of security for machine learning," *arXiv preprint arXiv:1611.03814*, 2016.
[116] C. Szegedy *et al.*, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.
[117] N. Papernot, P. McDaniel, and I. Goodfellow, "Transferability in machine learning: From phenomena to black-box attacks," *arXiv preprint arXiv:1605.07277*, 2016.
[118] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI- driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[119] N. Papernot, P. McDaniel, A. Sinha, and M. Fredrikson, "Towards the science of security for machine learning," *arXiv preprint arXiv:1611.03814*, 2016.
[120] R. Kaur, D. Gabrijelcv, and T. Klobucar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[121] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[122] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[123] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021. [124] R. Kaur, D. Gabrijelcv, and T. Klobucar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, p. 101804, Sep. 2023.
[125] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," *J. Big Data*, vol. 11, no. 1, p. 105, Aug. 2024.
[126] M. S. Hossain, M. M. Islam, and M. A. Rahman, "Artificial intelligence in cybersecurity: A comprehensive review," *J. Inf. Secur. Appl.*, vol. 58, p. 102693, Apr. 2021.
[127] S. Z. K. Arshad, M. A. Khan, M. A. Khan, and M. A. Khan, "A comprehensive review on machine learning based intrusion detection systems for IoT networks," *J. Netw. Comput. Appl.*, vol. 196, p. 103234, Dec. 2021.