Artificial Intelligence in Cyber Security

Md Fazley Rafy , *Graduate Student Member, IEEE*

***Abstract*—The unprecedented pace of technology has been significantly influenced by the integration of Artificial Intelligence (AI). The ubiquity of AI spans various domains, garnering both criticism and acclaim. Its growing application presents both advantages and drawbacks in cybersecurity, as it becomes a standard component in the development and operational phases of contemporary technologies. This paper provides a comprehensive overview of AI utilization in cybersecurity, exploring its benefits, challenges, and potential negative impacts. In addition to that, it explores AI-based models that enhance or compromise security across various infrastructures and cyber networks. The paper critically examines the role of AI in developing cybersecurity applications, proposes strategies for leveraging emerging technologies to counteract AI-generated threats and vulnerabilities, and addresses the socio-economic repercussions of the involvement of AI in cybersecurity.**

***Index Terms*—Artificial Intelligence, Machine Learning, Deep Learning, Neural Networks, Expert Systems, Natural Language Processing, Threat Detection, Real-time Mitigation, Social Media Security, AI-driven Threat Intelligence, Predictive Analytics, Adversarial Attacks, Privacy Concerns, Ethical Considerations, Quantum Computing, Dynamic Threats, Data Quality, Real-time Processing, Ethical Biases in AI, Cyber Attack, Cyber Security**

I. Introduction

In light of technological advancements and the proliferation of automation systems, the 21st century is witnessing an increasing inclination towards Artificial Intelligence (AI) across diverse application domains. Factors such as the dramatic expansion in computing capabilities, the ubiquitous integration of smart devices, and the swift evolution of businesses via datafication underscore the potential of AI-centric implementations. However, as we navigate this era of AI-driven innovations, concerns regarding the security and well-being of human society rise concomitantly. Considering the breadth of AI-centric applications, this paper is narrows its focus on the vulnerability, opportunity, and growth of AI in Cyber Security domain. The definition and motive of cyber security varies among diverse organizations and standards around the world. According to the ISO/IEC 27032:2023(en) Cybersecurity—Guidelines for Internet security, Cyber Security refers to the practice of safeguarding internet-connected systems, including hardware, software, and data, from cyberattacks, damage, or unauthorized access. It encompasses a range of processes, technologies, and controls designed to protect systems, networks, and data from cyber threats [1]. In such a case, safeguarding means keeping cyber risks at a tolerable level and protecting as many assets as possible. Incorporating cybersecurity is not to only secure the perimeters, but rather to identify the value of the assets under threat, localize the threat, and then prioritize the actions to build a defense-in-depth framework that ensures continuity of service [2]. In recent years, Cyber Attacks have targeted critical infrastructures, such as water supply, petrochemical installations, nuclear power plants, and transport infrastructure systems, to provoke power outages and compromise sensitive data [3]. In addition to rising threats, the convergence of IT and OT has increased the intricacy of the technical system due to the integration of physical devices and networked software and sensors. IEC recommends building cyber resilience from such intricacy and vulnerability through holistic approaches involving processes, people, and technologies, like AI [2]. However, AI-driven technologies have both positive and negative implications when incorporated with cyber security. While a part of AI buttresses the cyber resiliency, a counterpart trusses the infiltration. Research on AI-based applications is vastly diverse, though the security concern is far greater than precautions measures. The survey on cyber defense provided insights into the expectation of utilizing AI to automate the system to analyze the vast amount of data and perform critical inspections to identify cyber threats [4]. The study advocated for the neural network models to strengthen cyber resilience in military applications, and yet after a decade, the researchers are still exploring the opportunities and threats of using automated AI. The rest of the paper is prepared as follows: Section II discusses how AI-driven technologies evolved from the earlier novel stages to the modern current stages in the field of cyber security over the years. Section III includes the details of modern AI techniques and cybersecurity applications. Following that the challenges with the AI-integrated system for security measures are discussed in Section V. The threats of the same technology in the cybersecurity domain are also mentioned in Section VI. The paper concludes in Section IX following the future prospect of AI in the cybersecurity domain in Section VIII.

II. Evolution of AI in Cybersecurity

Growing technology allows the rapid growth in AI evolution in the 21st Century. However, different decades of research contributed to the current expeditious AI development. The groundwork was laid by Alan Turin in the mid-20th century by the introduction of the Turing test to evaluate intelligence in machines [5], and later furnished by John McCarthy to introduce the concept of generality in AI [6]. Though it was later in the century when conceptual theories became tangible implementations, these developments have continued to steer forward. During the 1990s, the world of technology encountered increasing computation power and promising development in data generation and processing systems. The

Md Fazley Rafy is with the LCSEE, WVU, doing his Ph.D. in Computer Science. Corresponding author: Md Fazley Rafy (E-mail: mr00065@mix.wvu.edu)

concept of Machine learning started to become prevalent and prominent in every other field of applications. The development of the internet with an exponential increase in computational capability in processing technology illustrated the true potential of Neural Network (NN) architecture with its capability of complex analysis [7]. During the same time in the late 1900s, the U.S. military urged the need for AI, to secure National Information Infrastructure (NII). Though the paper focused on the border security perspective of the spectrum of threats ranging from internal to transnational attacks, and the interconnected nature of military, government, and civilian information systems, the emphasis on improved measures to cope with the constantly evolving technology is crucial for maintaining the relevance and effectiveness of NII. AI was recommended for network intrusion detection and managing the cyber attack impacts as it can assist in analyzing vast amounts of data to detect attack patterns, and attack signatures, inform network intrusion tools, and improve decision-making capabilities. At that earlier stage of AI, the National Infrastructure Protection Center (NIPC) planned to integrate AI-based software developed by Sandia National Laboratories (SNL) into the Arms Control Treaty Monitoring System (ACTMS) where the developed embedded system was modeled with Intelligent Agent Sensors to detect attacks in critical infrastructures [8]. Entering the 21st century, progress took an appreciable shift from a steady state to a rapid sprint with the emergence of Big Data and sophisticated AI models. Improved algorithms, alongside the evolution of hardware capabilities like Graphics Processing Units (GPU) [9], enabled AI to process and learn from this data deluge with unprecedented efficiency. The foundational theories of AI were often motivated by statistical methods in that time domain, as those methods demonstrated decent performance with security event management tools for network monitoring and anomaly prevention. Building anomaly network intrusion using packet filtering, pattern matching, and flagging logs of events, the statistical models were utilized to develop intrusion and prevention mechanisms by observing significant deviations from normal or expected behaviors. Hierarchical Intrusion Detection System (HIDE) and the Generalized Anomaly and Fault Threshold system (GAFT)) were proposed to motivate statistical preprocessing and neural network classifications to detect network threats and faults [10]. The model used in HIDE consists of activity profiles generated and classified by several probability density functions. These profiles in turn are characterized to determine normal behavior in the monitored parameters. With its significant possibilities, there were still some limitations with statistical models. When dealing with cybersecurity data, there are numerous nonstationary network traffic involved in the system monitoring mechanisms. And since the network traffic patterns can alter over time that may become challenging for statistical models that depend on historical data to perform predictive analysis or anomaly detection. These functionalities of working on known patterns and historical data can also contribute to inefficiencies in detecting unprecedented types of attacks, commonly referred to as zero-day attacks. Moreover, the implementation of sophisticated models can require extensive computational power that may result in resource-intensive operations in large and complex network environments. Limitations of such statistical models can be resolved with more advanced network models governed by AI technology to better deal with the uncertainties and incompleteness in malicious data and unpredictable patterns in cyber attacks. Bayesian networks initially started to adapt to such complexities which were addressed by the Bayesian ontologies to enhance cybersecurity measures [11]. Bayesian ontologies and probabilistic reasoning from the Bayesian network can address cybersecurity challenges by achieving semantic interoperability among different systems and enhancing knowledge sharing. As evolving cybersecurity data can be significantly noisy and irregular, the uncertainty can be represented by ontologies in a principled manner. The probabilistic ontologies further build up standard ontologies by presenting constructs to relate statistical regularities and probabilistic interrelationships in a domain of application.The integration of such probabilistic constructs with Web Ontology Language (OWL) is proposed to highlight the balance between expressiveness and traceability in representing those uncertainties [11]. In cybersecurity, identifying and analyzing threats often involves interpreting patterns and anomalies in data. Probabilistic ontologies can improve this process by providing a framework for understanding the likelihood of certain events or behaviors, which is crucial for detecting potential security breaches or attacks. Capturing and comprehending the uncertainties for security analysis is essential for cyber security. As traditional graph models have limitations in supporting network defense mechanisms because of their limitation to address unknown vulnerabilities in real-time, the Bayesian network was utilized to overcome these challenges [12]. The graphical representation of the cause-and-effect relationship between events in the Bayesian network was illustrated as effective for uncertainties including unprecedented attack patterns, imperfect sequence of exploits, and limitations of anomaly detection sensors. The authors illustrated this model can amount to the causal relationships in cyber attacks and conditional probability tables (CPTs) to capture the likelihood of known and unknown attack scenarios. While the paper suggests that the Bayesian Network model should not be too sensitive to perturbations in parameters, but in practice, the Bayesian network highly relies on conditional probabilities to form the causal relationship between events. Small changes in those probabilities can significantly perturb the parameters and the outcome of the predictability. This sensitivity can be a drawback in a rapidly evolving field like cybersecurity, where the threat landscape changes quickly. To defend against these dynamic changes Adversary Courses of Action (ACoA) based on classical AI planning was proposed which is a sequence of exploits mapped from the initial state to the final action consisting of goals of an attacker [13]. The paper discusses the Behavioral Adversary Modeling System (BAMS) to simulate possible actions of insider threat using Hoffmann's METRIC-FF planner and Planning Domain

Definition Language (PDDL) to model a simplified Document Management System (DMS) and generate Courses of Action (COA). This method is then able to simulate the prompt generation of intricate threat models that can assist operators in predicting vulnerabilities and malicious acts of insider threats. During the early decade of the 21st century, rule-based systems were predominant in detecting anomaly patterns related to cyber threats. Intelligent Threat Prevention and Sensing Engine (ITPSE) architecture was proposed during that time, which supported different rule bases generated from expert knowledge systems and live traffic monitoring. The framework consisted of two engines; the Intelligent Threat Sensing Engine (ITSE) and the Intelligent Prevention Engine (IPE) [14]. The authors also discussed the need for an Intelligent Assistant System (IAS) architecture that includes a multidisciplinary approach of integrating traditional statistical models with AI techniques like data mining, fuzzy logic, and neural networks for better efficiency in security analysis. AI agents were introduced for advanced World Wide Web (WWW) interface computing to improve the interactions among heterogeneous computing environments [15]. The use of cybersignatures in such mobile computing allowed enhanced authentication and security. The authors further discussed innovative authentication techniques with intelligent trees, and agent language processing to develop robust methods for web interactions. Emphasizing the temporal and spatial information of network traffic, Genetic Algorithm (GA) based AI rules were generated to detect network intrusions with AI-rule integrated IDS [16]. In this intrusion detection, different processes like the translation of network relations and traffic patterns into the rules with the help of GA were introduced relating the stages of chromosome evolution through selection, recombination, and mutation. As the anti-virus scanner becomes less effective in detecting diverse and voluminous malware, such as viruses, Trojan horses, and worms, a framework was proposed to facilitate malware behavior threat analysis by clustering known and unknown malware [17]. The proposed method uses a machine learning-based Vector space model for efficient computation to assign vectors as behavioral patterns. The framework monitors system calls and actions and uses a Malware Instruction Set to encode those calls and arguments using numeric identifiers. Though the mechanism does not completely eliminate the malware threats, it enhances the capability to characterize and mitigate novel malware developments efficiently. AI advanced to a mature and specified prevention mechanism since the 2010s by assigning more proactive approaches than reactive measures taken in the earlier decades. Initially dependent on the rule-based practices in cyber security with AI, the modern techniques took a shift from traditional machine learning models to more dynamic and adaptive AI learning strategies to provide cybersecurity solutions. Meanwhile, modern AI systems can analyze and operate with vast amounts of data in real-time to offer more comprehensive security coverage compared to the scarcity in scale and complexity in earlier models. However, as AI technologies become much more proactive and efficient, the adversarial AI started to evade detection or generate sophisticated attacks as explained further in Section VI. To classify traffic data and detect system log anomalies, a Transformer model-based detection mechanism is proposed and shown to be a better performer than contemporary RNN and LSTM models [18]. While LSTM or RNN-based models can analyze long sequential data patterns, the complex relationship among the long patterns can be more effectively realized and analyzed by Transformer models. The self-attention mechanism also reduces the computation cost and enhances the complex anomaly patterns by capturing the relationship in the patterns [18].

III. Modern Core AI Techniques in Cybersecurity

AI, as a broad domain with cutting-edge technology, has different subcategories that are molded to facilitate cybersecurity applications. Machine Learning (ML) algorithms, Deep Learning (DL), NN, Expert Systems (ExS), and Natural Language Processing (NLP) are some of the specific models of AI. The use of these AI-based algorithms can address significant challenges associated with unprecedented and polymorphic attacks in Cyber Space (CyS) a lot more efficiently than the ones taken in conventional methods. Yet, to truly grasp their value and impact, it is essential to explore the underlying mechanisms and operations that power these techniques to get a better insight into their application in CS.

A. Machine Learning Models

Without being specifically programmed, Machine learning models enable machines to provide decisions or predictions based on the mathematical representation of curated data. Preprocessing the data before feeding it to the ML is the very first step to empowering artificial intelligence in cybersecurity. The mechanism is often interpretable and generates insights into the data by feature importance or ranking the most important features. To make suitable and accurate decisions to detect anomalous behavior, comprehending model decisions is crucial in these methods. Feature engineering follows suit after the processed data and then trained data assists in detecting anomalies or outliers in the malicious patterns [19]. The simple form of ML makes it computationally less costly than the DL models, and hence lack proper performance to analyze complex patterns and long sequence of data to identify anomalies. However, DL models are commonly known for their intricacy and lack of interpretability, which can make it challenging to explain their decisions. ML models can excel in such applications where transparency is indispensable [20]. To defend against Distributed Denial-of-Service (DDOS) attacks on the emerging Internet of Things (IoT) devices, a robust intrusion detection system, IntruDTree, was proposed with ML-based models [21]. The paper emphasized the use of ML, especially tree-based models that can learn from the security datasets preprocessed from IoT devices. The use of IntriDTree enhances the IDS decision-making process and reduces computational complexity by leveraging smaller feature dimensions and ranking essential features based on their importance. Compared to other traditional ML models,

tree-based models are shown to be more effective than generalized intrusion detection models. However, with the emerging volume of network traffic, and data communication due to rigorously integrated computing devices with communication capabilities, ML models lag behind DL in terms of dynamic detection mechanism; hence the technological shift towards the DL models to enhance network monitoring and cyber security [22].

B. Deep Learning and Neural Networks

The basic structure of deep learning consists of an input layer, a hidden layer or layers, and an output layer. Depending on the computational layers' arrangement and interconnection, several potential models were developed over the years [20], [23], [24]. The core idea behind the deep learning models is the critical comprehension of neural networks in the human brain. How different neurons interact with each other to make decisions proactively to known and unknown events, is the essential functionality in modeling the deep networks. Hidden layers in artificial neural networks (ANN) can consist of multiple layers to train the network to adapt to complex patterns and provide efficient analysis based on relative data. The most common architecture of deep learning, Convolutional Neural Network (CNN), gained widespread recognition for its efficiency and brought unprecedented achievement through the concept of deep learning. The limitations of the CNN were realized and mitigated by the Recurrent Neural Network (RNN) which is able to regain the information lost in the hidden layers due to the concept called gradient descent [25] with the help of loops in the network. The most cutting-edge technology with the concept of RNN is the Long Short-Term Memory (LSTM) networks and Autoencoders, which have cells to regulate the flow of information and decoder to reconstruct the output from the compressed representation of the input respectively [26], [27]. Rather than sequential processing of RNN and LSTM, Transformer models use the self-attention mechanism to capture long-range dependencies in data [28], which is highly efficient in analyzing sequential yet complex data patterns in network traffic to detect sophisticated cyber threats. These different deep learning models are used to build proficient anomaly and IDS to identify intricate patterns and anomalous behaviors. The models are capable of learning from large datasets to comprehend potential security breaches, classify the types of threats, and identify both known and unknown anomalies. In the environment of a Software Defined Network (SDN), a flow-based IDS was proposed to process network packet information using DNN [29]. SDN allows global network overview and dynamic updating of forwarding rules to support a robust detection mechanism for even unknown threats. With the increasing integration of computing devices, such as Electronic Control Units (ECU), in safety-critical vehicles, Controller Area Networks (CAN) are a major security concern in in-vehicle networks. To address the vulnerability in the ECU-ECU message broadcast, an unsupervised deep belief network (DBN) based IDS was proposed to enhance the detection ratio in CAN. Compared to traditional ANN-based IDS which utilizes predefined attack patterns or specification-based approaches to identify anomalies, the proposed DBN was shown more effective with unknown patterns during an attack by classifying statistical patterns and mapping complex non-linear relations. Traditional ML often struggles with the mutation in big data analysis whereas DL models are capable of extracting minor perturbations from complicated features. Emerging IoT devices often encounter small mutations in attack patterns that can be vulnerable to the associated system, but the use of DL can enable capturing covert malicious patterns for their self-learning and compression abilities [30]. In network traffic, the data is usually sequential in nature, hence the use of RNN can enhance the capability of IDS because of the model's ability to remember past units that can help in historical data analysis for more informed decisions [31]. Moreover, RNN can capture temporal dependencies and patterns in sequential data to better deal with high-dimensional features to classify specific patterns of cyber attacks [31]. Considering the advantage of managing long-term dependencies in data, LSTM-based IDS can maintain well-balanced high detection rates and minimal false alarm rates [32]. For a very long sequence of data, RNN can face a vanishing gradient problem during training that results in a poor detection mechanism. On the contrary, a variant of RNN, LSTM processes the long sequence with the in-build gating mechanism that allows for retaining valuable information from the long sequence of data during training and counteracts the limitation of RNN [33].

C. Expert Systems and their relevance

Being a subdomain of AI, Expert Systems are computer-based systems that emulate problem-solving and automated decision-making processes based on the combination of expert human knowledge-based systems and rule-based inference engines. The advisory capabilities of Expert System amplify the resiliency and robustness of the system in the context of cybersecurity [34]. Rule-based IDS are a common example of this concept where predefined rules are set in the detection process to monitor the abnormal behaviors in the network or system. As soon as traffic matches a rule or set of rules, the operators are informed of the anomaly to make prompt decisions. These rules or heuristics help to identify the traces of suspicious patterns in both system logs to provide offline security service and in live traffic monitoring systems to alert for anomalies. Rules and decision-tree-based IDS, RDTIDS, was proposed to classify malicious and benign network traffic with the embedded rule-based expert system and decision-tree approach [35]. This hybrid detection mechanism uses a three-tier fog computing framework, decision tree modeling with REP Tree and a rule-based classifier with JRIP that composes the first two parallel processes and feeds to the final classifier built with Forest PA to refine the findings which results in a more accurate representation of complex data patterns. Another comprehensive rule-based IDS with Java ExS Shell (JESS) that provides rules for the Pattern-based Intrusion Detection Engine (PIDE) is proposed to analyze user

behavior in the system [36]. The proposed PIDE acts as an ExS to identify suspicious behavior based on a defined set of rules as a pattern recognition engine. Integrating the DL-based classifier into the Rule-based feature selection method, a hybrid expert system model can achieve higher performance in detection rate and lower false positive ratio (FPR) [37]. In industrial IoT applications, ExS can prove to be beneficial in cybersecurity management by providing efficient monitoring of the vast amount of traffic. In the proposed hybrid architecture, the rule-based feature selection significantly reduces the subset of trained features and improves the DL-based classifier performance [37]. ExS plays a crucial role in cybersecurity by leveraging expert knowledge and rule-based reasoning to enhance security operations, threat detection, incident response, and compliance management. They serve as valuable tools for organizations striving to maintain robust and proactive cybersecurity measures.

D. Natural Language Processing

Abundant information flow online in social media and web pages motivated the analysis and interpretation of unstructured textual data to get concurrent trends in cyber threats and threats from social media sentiments. NLP models aim to extract and isolate those threat reports and malware code inserts from textual analysis. It helps to parse and extract relevant information to aid the identification of incidents from security logs [38]. Identifying phishing emails are widespread concern to avoid personal information leakage, like bank account details, social security numbers, and user passwords. Contemporary ML models lack accuracy in this domain due to the reliance on manual detection of representative features and DL models face similar challenges due to the deficiency of embedded words in the model for proper content representation in an email conversation. NLP can be integrated with this ML and DL models for efficient classification of email contents to efficiently identify phishing attacks [39]. Based on the cognitive analysis by ML models, NLP was used to create domain ontologies using a two-fold approach: symmetry stage and adjusted machine stage [40]. Based on the Ontologies, a prototype, defined as cybersecurity analyzer, was generated to obtain a detailed architecture with essential components and an implementation initializer. The structure is formed in four stages; first the the input as a document, which goes through the ML-based cognitive analysis, followed by data storing and visualizations to finally REST APIs to integrate all these functions. The cyber analyzer can efficiently analyze cyber documents for forensics and malicious entity identification using NLP methods. Due to the colossal volume of web content and open-source texts, security analysts regularly confront obstacles to discovering cyber threat-related content online within optimal time. To address this issue, an automated system to extract threatening cyber content from publicly available online data is proposed using a naturally embedded method, referred to as Doc2Vec [41]. The embedded method uses a natural language processor as a filter to isolate the cybersecurity-specific contents from feed data. NLP reduces the need for guided intervention in the midst of regular operations by analyzing unstructured data like security content, operation logs, and threat intelligence and providing valuable insight to the cybersecurity specialist to perform preventive actions. A study was performed to list the different attack patterns in the Common Attack Pattern Enumeration and Classification (CAPEC) database to assist cybersecurity experts in analyzing the suggested attack events for proactive measures [42]. The study utilizes topic modeling, in which several unstructured topics are grouped to make a single comprehensible structure, to bring out hidden vulnerable information from the attack description in the database. Different stages of the filtering process involve generating a topic model, creating a term-frequency vector, and estimating the posterior distribution of topics before evaluating the KL divergence between the topic distribution of the corresponding system and the corresponding attacks. The entire process thus ensures the enriched performance of practical cyber security threat detection and assessment. CyberAttack Sensing and Information Extraction (CASIE) was proposed to further enhance the applicability of NLP in cyber threat identification from textual forms as the system was modeled to keep the general public informed of the cyber events through knowledge-based graph representation in different online articles [43]. CASIE was trained on a tremendous amount of news articles to label several vulnerable events, like Databreach, Phishing, Ransomware, Discovering Vulnerability, and Patch Vulnerability, including their semantic roles and several event-relevant argument classes. Along with the beneficial applications, NLP's models can lack accuracy in performing cyber threat analysis because of their dependence on the quality and representativeness of the training data, which can be a limiting factor.

IV. Opportunities and Advancements

The widespread adoption of AI techniques across various applications holds tremendous potential for addressing numerous socio-economic and environmental challenges. However, to fully unlock this potential, it is imperative to prioritize research efforts aimed at securing these technologies. One of the critical areas garnering substantial research attention is adversarial machine learning [44]. A sustained and dedicated focus in this domain is crucial as it plays a pivotal role in ensuring the reliable proliferation of transformative AI technologies. As AI continues its integration into diverse facets of human activities and daily life, the development of robust algorithms takes on paramount importance. These robust algorithms are not merely desirable but absolutely essential for fostering a future characterized by innovation, safety, and security. Thus, nurturing advancements in adversarial machine learning will fortify AI's foundation, enabling its responsible and secure deployment to tackle pressing global challenges while safeguarding individuals and organizations from potential risks. AI has made remarkable notes in improving cybersecurity applications through its contributions to threat detection, response, and prevention. Machine learning and deep learning algorithms are leveraged to analyze exten-

sive datasets, assisting in the identification of patterns and anomalies for timely threat detection. Furthermore, AI-enabled systems improve the routine security tasks of the automation industry, enabling cybersecurity operators the opportunity to dedicate their expertise to more interactive challenges. Powerful security measures are also implemented through AI-powered authentication methods, such as biometrics and behavioral analysis [45], [46], which effectively defend against unauthorized access. Moreover, AI exhibits the capability to adapt rigorously to the ever-evolving threat landscape, ensuring proactive actions in cybersecurity defense methods improvisation, collectively reinforcing cybersecurity protocols, and safeguarding critical data and infrastructure within the context of our increasingly interconnected digital world. Furthermore, the influence of artificial intelligence (AI) on the field of cybersecurity extends far beyond its traditional role in threat detection as AI empowers organizations to conduct predictive analysis, effectively identifying vulnerabilities and potential attack vectors [47]. This proactive approach enables organizations to address security weaknesses preventively. Additionally, AI-driven technologies enhance incident response, enabling organizations to swiftly mitigate the impact of cyberattacks. The contribution of AI is not limited to threat detection; it also plays a crucial role in user behavior analysis, facilitating the identification of insider threats and unauthorized activities within organizational networks [48]. The capacity for large-scale data processing and the generation of actionable insights has ushered in a transformative era for cybersecurity. This technological advancement enhances adaptability and proficiency in defending against increasingly sophisticated cyber threats. As AI continues to advance, its role in safeguarding digital assets and ensuring privacy will become ever more critical in today's interconnected digital landscape. Host-based and Network-based IDSs are the most widespread applications to address the cybersecurity concerns in almost every field of application. The following sections categorically outline the contributions of various AI techniques in enhancing cybersecurity measures, providing details on their formation and methods.

A. Anomaly Detection

Deviation from normal behavior, which is defined by the specific organizations and resiliency requirements, is identified and mitigated through the utilization of diverse statistical models. The pattern recognition of the attack vectors through these models helps the security personnel to make efficient decisions in a timely manner. However, often the model is used in amalgamation of other methods to realize the true nature of the threat and address the proactive analysis. A hybrid model with a Gaussian Mixture was proposed for anomaly detection with the integration of misuse of network detection using the decision tree model [49]. The model takes account of the inconsistencies in the dataset or logs that can lead to misguided, or even worse, high false rates in the detection mechanism. The decision tree facilitates the conditional statement-based rules to separate the anomalous pattern from the normal ones. The best attributes from the decision tree model are used to make the final decision on choosing the usual pattern and isolating the pre-defined attack patterns. In the decision tree, each normal leaf is modeled using Gaussian Mixture Models (GMM). This approach posits that observations originate from multiple Gaussian distributions, the parameters of which are not predetermined. These parameters are estimated through the Expectation Maximization (EM) algorithm. This method is notably efficient in identifying attacks that bear resemblance to normal distribution patterns. The hybrid model classifies the normal pattern of data into multiple subsets to better comprehend and analyze the normal pattern to identify attack patterns. The inclusion of IoT devices in smart grid systems sometimes bears the responsibility of making the Advanced Metering Infrastructure (AMI) vulnerable. Where traditional methods of securing the AMI fail to address the vulnerability properly, the GMM-based detection mechanism can improve the system by addressing the reliance on external knowledge and a predefined threshold [50]. The capability of GMM to cluster historical data to define different ranges of normal patterns enhances the detection of malicious events without dependence on external inputs and with robustness against data fluctuations. Compared to the Hidden Markov Models (HMM), GMM does not require specific training data or predefined knowledge and thus is highly robust against frequent data fluctuations to identify anomalous data [51]. To develop a smart city with lots of ICT devices, IoT communications, and cloud-connected data storage systems, AI-integrated models can be embedded at various levels of the system to enhance anomaly detection. A hybrid model with centralized to distributed architecture represents the necessity of AI-based ML models to secure edge-to-cloud networks in smart cities [52]. The architecture addresses the attacks on distributed computing devices with the facilitated ML-integrated SDN network with cloud-based services. The cooperation of SDN, multiple controllers, and ML methods at the edge networks ensures better security from malicious or abnormal data and identifies corrupted system resources. Often cooperation of expert systems extends the performance of the detector by feedback on the accuracy and detection results from system analysts, and operators, enabling the AI model to learn from the human in the loop system. An ensemble tree-based technique with an Isolation Forest anomaly detector was proposed with the ExS for distributed system [53]. The model uses expert feedback to enhance the understanding of the attack patterns and the detection process by reducing false positives and increasing accurate attack pattern identification. External flow of data for instance, in content delivery networks, the AI-integrated models (including both Isolation Forest and GMM) are efficient as well in detecting attacks such as DDoS and Cache Pollution Attack (CPA) [54]. While ML and statistical methods with expert systems have widespread applications in cybersecurity, NN-based LSTM model is also prevalent in numerous anomaly tool detection and attack prevention methods [55]–[57].

B. Signature-based Detection

To detect more salient features in the data pattern with time-dependency and time-urgency, the use of signature as an attribute for the detection mechanism is quite popular in the literature [58], [59]. Subdivided into two mechanisms, signature-based models work with either the ruleset that defends against unknown or undefined activities in the network or with the patterns to isolate the abnormal behavior instances from the normal traffic [60]. In rule-based models, the predefined set of rules dictates the filtering of attack and normal patterns in data, where the most popular tools for such detection are Snort, Suricata, Zeek, etc [61]. Apart from these open-source IDS tools, there are multiple other rule-based methods developed, one of which is a Fuzzy rule-based model for hazard identification in cross-country product pipeline system [62]. The model utilizes a Fuzzy Rule Base (FRB) approach with Grey Relations Theory (GRT) to address imprecise, uncertain, or subjective information to implement more nuanced and adaptable risk analysis in a complex and unpredictable environment like in a pipeline system with limited data. Increasing phishing, business email exploitation, and ransomware have been noticeable during pandemics. Rule-based Fuzzy Logic and Data Mining techniques are used to minimize the impact of potential adversaries and threats [63]. In pattern-matching methods, the deviation from the normal and historical data is measured, and the ability to detect even the slightest deviation from properly addressing the pattern type is challenging yet has been done with different models, like Absolute Median Deviation (AMD) for refined detection and prevention process [58]. In the historical data, known attack patterns, unharmful events, and Logs of different categories of events are documented and processed with a pattern recognition mechanism to make effective decisions on the real-time unknown data. Pattern-matching methods are also integrated with multiple other techniques to incorporate the required analysis and enhance the capability of detection, as described in the development of Digital forensics face pattern recognition using PCA, NN, and GA integrated model [64]. After dimensionality reduction by PCA, NN, and GA are utilized for optimized pattern matching in this federated model.

C. Cloud Security and Encryption

Cloud-based data communication within infrastructure, different applications, and databases requires utmost importance in terms of secure authentication, resource access control, and privacy protection. An energy-efficient industry-scale data management in an IoT environment is proposed utilizing EEIBDM framework [65]. This framework enables a digital twin system with a virtual representation of Industrial IoT-based big data management and combines reinforcement techniques and federated learning to strengthen cloud security. By employing DNA-based Huffman coding in the encryption process, a more complex and less predictable method of encoding data was inspired by biological resilience [66]. The proposed method greatly improves the security and robustness of the cloud by encrypting data with DNA-motivated Huffman coding rather than traditional binary coding. Internet of Vehicles (IoV) is designed to have interconnected entities, like sensors, and traffic management systems, which are highly vulnerable to data transmission attacks if not secured properly. An efficient encryption algorithm is proposed, EAST, with steganography to conceal the encrypted data to provide multi-layer AI-driven protection from such cyber attacks [67].

D. Incident Response & Mitigation

As the detection and security monitoring techniques assist in identifying the adverse events, the subsequent actions require the analysis of the encountered threat and responding to the vulnerability. AI has not only contributed to the detection methods but also to the prevention and attack response mechanisms in modern technological infrastructures. Increasing interest and development of automated infrastructures involve the security concerns that are effectively addressed by AI models like statistical decision trees [49] to reach the best optimal decision, or NN models to automatically resolve anomalous events after detection [56]. AI can also enhance endpoint security by real-time monitoring and assessing events from endpoint devices for potential malicious activities or detected infiltration [68]. Such a prevention mechanism is able to assist in the timely identification and neutralization of vulnerabilities at the device level, mitigating the impact of the threat across the whole network. In wireless sensor network deep learning models are leveraged to identify potential threats in the network followed by the CNN based prevention mechanism to almost perfectly preventing the sensor attacks [69]. Phishing attacks consisting of interaction with malware software or social engineering causes great deal of damage in infiltrating the network and system. To prevent such adversity, the provision of AI driven training and assessments to the operators can yield massive improvement in prevention mechanism [70]. As the modern attackers are increasingly using AI to their benefits, the contermeasure would be as effective as with the inclusion of the similar AI technology to defend against sophisticated attacks. Along with these preventive analytics, the AI technology assist in adaptive decision making process to make a more interactive prevention mechanism to adapt with the ever changing attack vectors and strategies. An intelligent AI based Decision Support System was proposed to enforce risk management decision support for critical industrial infrastructures. The decision model encompasses AI based strategies such as multi-criteria analysis, causal networks, , and knowledge engineering to support complex industrial installations. Sensor and historical data based model helps to facilitate threat prevention, protection, and mitigation for potential multifaceted cybersecurity concerns [71].

E. Threat Intelligence & Deception Technology

Advancing from the active protection mechanism to more proactive methods, modern security management systems often allow vulnerable spots within specific parts of the infrastructure or tracking mechanisms within the system to capture the behavior or data flow from cybercriminals. The

method of luring the adversaries is often referred to as "Honey Pot", which assists the system analysts in gathering important information on the attack signatures and tactics of attackers. The analysis from Honey Pot can further be used to model the IDS or even to divert the attack from essential targets [72]. The data occupied by the Honeypot mechanism is a rich source of real-time threat intelligence. By analyzing this data, security analysts and forensic specialists can isolate new attack patterns, malware injection techniques, and exploitation trends [73]. As security measures need continuous updates to keep up with the modular improving attacks, threat intelligence is crucial for ensuring cyber resilience and staying ahead of attackers. In [73], the proposed AttackKG system automatically extracts and analyzes attack patterns from unstructured Cyber Threat Intelligence (CTI) reports. The analysis is redirected to the knowledge graphs to create templates and a revised graph alignment algorithm to modify the identification techniques in the system. The ability of AI in such an environment with a large volume of unstructured data improves the accuracy of threat detection, reduces manual interventions, and exemplifies the prospect of AI in preventing sophisticated cyber attacks.

V. Challenges in Implementing AI

While the opportunities are yet to be realized entirely, AI technologies are becoming more and more part of the developing applications, algorithms, and even human society. The progress is often hindered by certain challenges and limitations of AI in cyber security. One of the major concern is that the models and strategies discussed and even beyond the discussion in this paper depends highly on the quality of data and availability of massive amount of data to learn representation of the optimal patterns. Without proper knowledge of the system through massive and quality data, AI can lack in performance to defend against unprecedented cyber threats. Understanding these limitations are cruital for associating efficient operation of AI. There are several recent techniques to work on smaller dimension and scale of data to learn the patterns, although the diversity of data is significant to accurate predictive analytics or threat prevention or detection mechanism. Algorithms, such as, Transfer learning, Data Augmentation can successfully train on specific data patterns to assist in the detection process, yet, their lack of training on diverse patterns can undermine the performance in real time [74], [75]. Few Shot learning, active learning methods can work on real-time data with small datasets, but again needs a lot of diversity in the data to properly address real world attack events with continuous modification in the strategies to infiltrate and damage the normal operation [76], [77]. In the proposed few shot based anomaly detection model [76], category-agnostic is associated to aggregate data from different categories. The model compares the normal registrated features with the new ones to find dissimilarities in the image data with the help of statistical distribution estimator. While the model intends to generalize the categories with few short learning, it might be infeasible to the real world novel data patterns or different unseen category, which is usual for sophisticated attacks. Considering the required abundant data for accurate decision making and results, the large amount of data storage are also vulnerable to privacy breach and security risks. Frequent monitoring of the base raw data which is utilized to provide dynamic detection and prevention strategies in the system is computationally expensive and requires utmost security and robust management. It is reported that in 2021, with botnet, more than 80 million data breaches tool place and the number of that is increasing in exponential rate till now [78]. Another hurdle for the operation of AI is in real-time system behaviour control and management. As cybersecurity measure usually demands prompt response and prevention, the minimal delay between detection and prevention using AI can cause serious damage to the system while urging for real-time defence to be effective [79]. Though Particle Swarm Optimization (PSO) was shown to demonstrate higher accuracy with lower speed of detection and prevention compared to the traditional advanced models like the genetic algorithms and ant colony optimization, the lack diversity in the trained data set is absent in the analysis to develop comprehensive response to the complexity of speed vs accuracy [80]. The compromise follows another challenge which is the ability to properly address the avolving threats in the modern era with increasing access to AI technologies like, Generative Pre-trained Transformer (GPT) [81], Bidirectional Encoder Representations from Transformers (BERT) [82], and abundant online resources on hacking [83] etc. Evolving nature of cyber attacks makes it quite difficult to the developers and challenges the capability of AI based methods to perform reliably against the unprecedented events or sophisticated strategies [84]. Unknown attacks, commonly referred to as Zero Day attacks, are the most notorious one to harm the respective system if not identified timely and mitigated efficiently, in which signature based AI detection methods lag in performance. Nowadays, the attackers are as subtle as possible to stay inactive in a network by only eavesdropping for even several months before performing the targeted attacks. These stealthy, sophisticated, long-term cyber attacks, known as Advanced Persistent Threat (APT), can evade traditional security measures quite easily [85]. The attackers in APT can take remote control of the devices and cause severe internet threat. These covert, advanced, and persistent nature of the attacks are able to generate massive destruction to the critical infrastructures and particularly difficult for traditional AI based methods to defend against as seen in multiple events, like, Struxnet, BlackEnergy Attacks in 2015 [86], notpetya attack in 2017 [87] and so on. A game theory approach was proposed to address multiple attackers with multiple defenders instead of traditional focus on single attacker-single defender using multi-agent deep reinforcement learning to address the APT [88]. However, the method is more complex and computationally expensive than the traditional counterpart, and also can yield at a slower rate for player learning with the increasing number of player or states. Ethical consideration and biases in AI, particularly in the context of cyber security, are multifaceted and complicated to comprehend. ML, DL,

and NN models have the black box property with lack of transparency in the interepretibility of the model that can be questionable and risky in safety critical system or in critical automated infrastructures. AI models uses ZIP code or other language attributed for predictive analysis, which can generate misinterpretations considering the biases present in the data the model is trained on. Ethical consideration in general requires the involvement of Human operators, although the use of AI in consistent decision making is a large concern without the intervention of human specialists, which leads to another concern of over-reliance on AI [89]. The continuous advancement of AI has also led to the misuse of technology, posing risks to organizational networks and critical systems. In many cases, the knowledge and capabilities of AI can be wielded as threats to cybersecurity rather than serving as a benefit. It is imperative to exercise effective monitoring, regulation, and control to ensure that AI remains a blessing rather than a curse in the domain of cybersecurity. Knowleged about those threat and vulnerabilities are also essential to maintain the adaptive and proactive defense mechanism to fight against the notorious portion of AI application to raise security concenrs.

VI. Threats and Vulnerabilities

AI applications bring significant advantages to the development and operation of secure networking and system monitoring, yet the inherent threats and vulnerabilities diminish the reputation of the technology with the growing interest in AI. The most infamous use of AI is the negative implementation of the technology by malicious actors to cause harm to the automated industry with the very same methods designed to protect the system. The modular characteristics of AI can be used to mold the operation into threat and destruction rather than safety and reliability. The vulnerabilities in the AI methods in addition to that raise much more security concerns and pose a threat to exploitation, where attackers can manipulate the algorithms, invoke abnormal behaviors in the mechanism, and launch attacks such as adversarial attacks. The legitimate purpose of AI in such a way is often diverted by cybercriminals to gain personal benefits. Advanced phishing attacks, automated hacking, sophisticated fraud, and manipulations are prime examples of AI targetted attacks with the help of AI [90]. AI-incorporated attacks can be challenging to the security of the system as it can adapt to the security measures to evade the detection, prevention, and mitigation techniques. Another angle to look at is that AI can be very much biased as it is still trained on a certain amount of data with limited distribution in categories. The bias can increase the vulnerable points in the system to attract the adversaries to make use of those vulnerabilities. To better design the AI technology to reshape the cyber security methods to tackle modern, complex, and sophisticated attacks, the following limitations need to be addressed properly and more research is indeed essential for improvements.

A. Adversarial Attacks Against AI Models

The most infamous way to use AI by malicious threats is to incorporate Adversarial attacks. The foundational block of such attacks is built on the concept of adversarial learning or deep learning models. Generally, ML and DL models function based on the assumptions provided through data and features to classify normal and abnormal behavior in the process. The adversarial models make use of that concept to manipulate those assumptions to change the motive of the model and cause personified ramifications [91]. The attack model with adversarial learning initiates with the alteration of confidentiality in the ML or DL model on which the security measures or applications are running. By acquiring the knowledge of the model, capturing the features of the learning, and eavesdropping on the system requirements, the threat actor gains significant knowledge to move with the next step of compromising integrity. The features or inputs that are used to train the models are altered according to the malicious purpose of the attacker. As the attacker gains knowledge of the system in this process then follows the compromisation of integrity. The expected input features or learning process is altered to change the desired outcome from the models and the security measures are then made incapable of identifying anomalies as these are modified to get passed the IDS or IPS during the attack vector with the help of adversarial attack. Subsequently, the attacker is able to change the availability of the services by changing the integrity of the system model. Normal behaviors are no longer recognized and malicious events are not prevented or recognized efficiently by the security applications. With this sophisticated attack, the threat actor is capable of compromising the CIA triad of the cyber security integrated system to damage the respective infrastructure or critical services [92]. These manipulations often exhibit subtlety and complexity, posing challenges in detection. The consequences of such actions are far-reaching, potentially resulting in the circumvention of security measures or the distortion of automated decision-making systems. To counteract these threats, it is imperative that AI models undergo continuous evaluation and refinement. This process should include adversarial training, wherein the models are exposed to potential attack scenarios during their developmental stages, thereby bolstering their robustness and capability to withstand such tactics. Though DL-based IDS can differentiate the subtle changes in the normal and abnormal events, the adversarial attack on these NIDS can bring challenges to the identification mechanism. In an IoT environment, the attack has been shown to be very effective in changing the efficiency of the security structure. With the use of adversarial attacks, small changes in the bytes in malicious packets were shown to effectively reduce the performance of the predictive model in the IoT-based NIDS [93].

B. Privacy Concerns

Privacy concerns arise prominently when AI tools process sensitive data for security purposes. The aggregation and analysis of large datasets can inadvertently expose personal

information, leading to privacy breaches. This concern is further amplified by the increasing sophistication of AI tools capable of deriving detailed inferences from aggregated data. DL models have widespread applications in the cybersecurity domain due to their comparable capability to extract important features without external feature engineering and their compatibility to produce better results with correlated data. The learning process of this technique includes large data sets, with the help of which the model learns to differentiate between different patterns. The predictions or classifications are performed based on the input data provided to these models. However, the concern grows when the model memorizes the core characteristics or details from the training process as in most of the cases the data can be private or sensitive. These models thus become a security concern since the attackers can target the confidentiality of these models and gain such sensitive and personal information [94]. For instance, if the model is trained on medical data or financial information, the data leakage can bring about identity theft, privacy violations, or financial fraud [78]. This vulnerability of data leakage can be exploited through reverse engineering or querying a model to unveil the critical information stored during the training stage. With a collaborative DL process, multiple parties can share a specific set of model parameters to train the available DL model and can get a flexible advantage from this collective learning process. However, due to the availability of the partial parameters in the model, it can be exploited by the Generative Adversarial Network (GAN) attacks. The GAN is a dual neural network model with a generator and discriminator-based network. The generator tries to create authentic data and the discriminator distinguishes the generated data from the original one. Though initially, the generator remains the weaker model, the distinguishing remarks of the discriminator continue to make it stronger and after a while, the generator creates indistinguishable data. The same has been performed with the collaborative DL learning model, where the GAN recreates private data from the shared parameters during the attack [94]. These attacks can generate similar data based on the evaluation of the discriminator and extract information without accessing the shared parameters directly. Even with the presence of the differential privacy measure, integrated to protect the shared parameters, the GAN attack can extract the information easily. Because the GAN model replicated the data with the help of partial parameters and did not exactly steal the information, the attack model was able to bypass the security measures. Therefore, implementing stringent data protection measures, such as data anonymization, encryption, and secure data storage and access protocols, is critical. Moreover, AI models must be designed to balance the need for security with the imperative of preserving user privacy.

C. Potential Misuse of AI

The potential misuse of AI technologies in orchestrating cyber-attacks presents a dual-threat landscape. AI can be exploited to create sophisticated malware that adapts to security defenses, automate large-scale phishing campaigns [95], and even conduct deepfake attacks that manipulate audio and visual content [96]. The increasing accessibility of AI tools and techniques means that the barrier to entry for cybercriminals is lowering, leading to more advanced and targeted cyber-attacks. It necessitates a proactive approach in cybersecurity, where defensive strategies are continuously evolving to counter AI-assisted threats.

VII. Socio-Economic Effect

It is important to consider that the improvements and progress we experience through any technological changes have a radical impact on the quality of human life and the nature of job responsibilities. Due to the multifaceted impact of AI in cybersecurity, the technological revolution is not the only aspect to get affected, but also the different aspects of society experiencing the flow of change. The following section will explore how the changes in societal interaction in the digital sphere, and how the transformation of social life is noticeable in the future due to the changes in technology. The impact also extends to the responsibilities of people in work, their motivation to continue skill-based jobs, and the risk of losing certain branches of the workforce because of the revolution. These insights on the societal and workforce changes are essential to have a comprehensive understanding of the implications of the role of AI in our modern digitized world.

A. Societal Impact

AI-driven technologies are continuously improving their mechanism to provide several benefits from sharing personal data. The progress instigates trust and reliability on the digital platform where the use of personal information can seem to be secure and beneficial in many aspects [97], [98]. However, measures to secure the use of these sensitive data are hardly always the most effective ones. These measures can often fail to protect the integrity of the individual who holds the AI-based technological platforms trustworthy [99], [100]. For instance, the smart home assistant devices pose a high-security risk to the breach of information and interruption of services with the open nature of voice channel, continuous dependence on AI support, and complex underlying architectures of the software system [100]. Similarly, increasing the use of chatGPT can indirectly support the performance of social engineering attacks, automated hacking, network payload generation for attacks, sophisticated malware development, email phishing attacks, and so on [99]. Another aspect of the growth of cybersecurity technology with the help the AI is that the resources to utilize the most advanced technology may not be available to all. Lack of resources, especially financial limitations, can cause digital separation among different organizations. The limitation may lead to the attention of attackers to perform sophisticated cyber attacks on such weaker networks and services. Ethical concern is also associated with every innovation in history; AI is not different from that perspective. Maintaining ethical use of AI requires consensus among nations, professionals, researchers, and even individuals. However, it will be very difficult to

obtain such equality in maintaining the same level of ethics for the use of AI among different sectors. Those dissimalirities in ethical approach may give rise to the targeted cyber attacks and security risks as depicted in the literature [101].

B. Cybersecurity Workforce

In the current technological landscape, the integration of AI into cybersecurity has profound implications for the workforce. The reliance of AI on high-quality data necessitates a workforce proficient in data-centric skills to analyze and interpret model behaviors effectively. This expertise is crucial for developing robust and secure cybersecurity applications. The advent of AI has shifted the focus from repetitive, manual tasks to roles that demand critical thinking and the ability to handle complex scenarios. This evolution is expected to continue, with AI models taking over more routine responsibilities. As a result, cybersecurity professionals will be able to concentrate more on strategic decision-making and the overall management of cybersecurity frameworks, rather than on labor-intensive tasks that require less critical thinking [102]. This transformation underscores a significant trend: the move towards a skill set centered around critical and strategic thinking. The cybersecurity workforce must adapt to this change, acquiring skills that cannot be replicated by AI. This includes not only technical capabilities but also a deep understanding of ethical, legal, and social implications of AI in cybersecurity. The impact of AI on routine jobs in the cybersecurity sector is a double-edged sword. While it may lead to a reduction in certain job roles, it also opens up opportunities for professionals to elevate their skills and engage in more meaningful and impactful work. The emphasis on multidisciplinary skills is becoming increasingly important, with cybersecurity professionals needing to balance technical knowledge with an understanding of ethical, legal, and social considerations.

VIII. Future Directions

The integration of AI into cybersecurity, while offering advanced defense mechanisms, also introduces complex challenges. The arms race between AI-powered security measures and AI-assisted adversarial strategies necessitates a dynamic and evolving approach to cybersecurity. Continuous research, ethical considerations, robust privacy protection measures, and the development of resilient AI models are essential in navigating this landscape. The future of cybersecurity lies in the ability to anticipate, adapt, and respond to these challenges effectively. One way to think is the association of emerging technologies with AI-based models, such as Quantum Computing. The prospect of this computation method can transform cybersecurity by potentially revolutionizing current cryptography methods. However, it also offers the development of quantum-resistant algorithms, ensuring a new level of data protection. Hence, the future prospects of AI in cybersecurity include improving quantum encryption techniques like Quantum Key Distribution (QKD), which could offer robust security levels against modern dynamic attacks [103].

Blockchain is another emerging technology that ensures integrity management for effective decision-making processes, especially successful for its security measures, transparency, and immutability to manage IoT networks [104]. Even though blockchain technology can not inherently gain AI-like learning capabilities, it can complement AI-based cybersecurity tools and applications to enhance data security and integrity from sophisticated attacks, like Adversarial Network attacks. Even the combination of Quantum computing and Blockchain can be a better prospect for AI-based security model improvements, as the Quantum-inspired technique was integrated with the blockchain framework for reliable data transmission [105]. Explainable Artificial Intelligence (XAI) plays a crucial role in the future direction of AI in cybersecurity. It addresses the challenge of the ”black box” nature of deep learning models, aiming to make AI’s decision-making transparent and understandable. This is vital in cybersecurity, where understanding the rationale behind AI-driven security decisions is essential for trust and reliability. Methods like interpretable models, post-hoc explanations, visualization tools, feature attribution, model simplification, natural language explanations, and prototypes and criticisms are instrumental in achieving XAI’s goals. Making AI models in cybersecurity explainable ensures not only trust and fairness but also enhances the efficacy of AI systems in critical security applications [106].

IX. Conclusion

The role of AI has been appreciated and challenged in different domains and applications. However, the domain of cybersecurity is the critical one that can not tolerate the failure of AI techniques as security is the only thing between threat actors and sensitive information and services. Hence the limitations and strengths of AI are essential for future researchers and developers to make proper adjustments and modifications to make proper decisions. This paper is prepared to address the unlimited possibilities of AI in the cyber cybersecurity field and at the same time realize the threats associated with it if the provided limitations are not considered during implementation. The evolution of AI is also provided to guide the researchers to understand the flow of changes that led to modern AI technology and relate the changes to bring the potentiality of AI to the forward direction as it has been done by numerous researchers, professionals, and individuals throughout history. The strength of AI can be maintained by the evolving cyberattacks if the evolution of AI can be run at the same rate as the attacks. Considering the possibility of integrating other technologies into the AI-integrated security applications, the threat actors can be defended effectively, or at least the impact can be minimized. Moreover, as for the shift in other domains as an impact, AI in cybersecurity is transforming the workforce, steering it towards more intellectually demanding roles and necessitating a diverse skill set, while also presenting opportunities for growth and adaptation in an evolving tech landscape.

X. Appendix A List of Acronym

The following is the list of Acronyms and their full forms used in this paper.

ACoA	—	Adversary Courses of Action
ACTMS	—	Arms Control Treaty
	—	Monitoring System
AMI	—	Advanced Metering Infrastructure
ANN	—	Artificial Neural Network
AI	—	Artificial Intelligence
BAMS	—	Behavioral Adversary Modeling
	—	System
CNN	—	Convolutional Neural Network
CPA	—	Cache Pollution Attack
CPT	—	Conditional Probability Tables
CR	—	Cyber Resilliency/ Cyber Resillience
CS	—	Cyber Security
CTI	—	Cyber Threat Intelligence
COA	—	Courses of Actions
DBN	—	Deep belief Network
DDOS	—-	Distributed Denial-of-Service
DMS	—	Document Management System
DL	—	Deep Learning
ExP	—	Expert Systems
FPR	—	False Positive Ratio
FRB	—	Fuzzy Rule Base
GA	—	Genetic Algorithm
GAN	—	Generative Adversarial Network
GAFT	—	Generalized Anomaly and Fault
	—	Threshold system
GMM	—	Gaussian Mixture Models
GPU	—	Graphics Processing Unit
GRT	—	Grey Relations Theory
HIDE	—	Hierarchical Intrusion
	—	Detection System
IAS	—	Intelligent Assistant System
IDS	—	Intrusion Detection System
ISO	—	International Organization for
	—	Standardization
IEC	—	International Electrotechnical
	—	Comission
ITSE	—	Intelligent Threat Sensing Engine
IPE	—	Intelligent Prevention Engine
IT	—	Information Technology
ML	—	Machine Learning
NLP	—	Natural Language Processing
NIDS	—	Network Inrusion Detection System
NII	—	National Information Infrastructure
NIPC	—	National Infrastructure Protection
	—	Center
OT	—	Operational Technology
OWL	—	Web Ontology Language
PDDL	—	Planning Domain Definition Language
QKD	—	Quantum Key Distribution
SDN	—	Software Defined Network
SNL	—	Sandia National Laboratories

References

[1] ISO/IEC. Cybersecurity — guidelines for internet security. https://standards.iteh.ai/catalog/standards/sist/2d12469a-69be-4365-88bb-05df3b0212db/iso-iec-27032-2023, 2023.

[2] International Electrotechnical Commission. Iec — cyber security. https://drive.google.com/file/d/1j0z2tmiajq5ff8ZfDPEwbHHIFXBwJIV5/view?usp=sharing, [2022].

[3] Tomas Plėta, Manuela Tvaronavičienė, Silvia Della Casa, and Konstantin Agafonov. Cyber-attacks to critical energy infrastructure and management issues: Overview of selected cases. *Insights into regional development. Vilnius: Entrepreneurship and Sustainability Center, 2020, vol. 2, no. 3.*, 2020.

[4] Enn Tyugu. Artificial intelligence in cyber defense. In *2011 3rd International conference on cyber conflict*, pages 1–11. IEEE, 2011.

[5] Alan M Turing. *Computing machinery and intelligence*. Springer, 2009.

[6] John McCarthy. Generality in artificial intelligence. *Communications of the ACM*, 30(12):1030–1035, 1987.

[7] Xavier Seuba. Big data, ai and border enforcement of intellectual property rights. *Intelligence*, 4:707–748, 1982.

[8] James H Thomas and ARMY WAR COLL CARLISLE BARRACKS PA. *Managing Risk to the National Information Infrastructure*. US Army War College, 1998.

[9] Jeff A Stuart and John D Owens. Multi-gpu mapreduce on gpu clusters. In *2011 IEEE International Parallel & Distributed Processing Symposium*, pages 1068–1079. IEEE, 2011.

[10] Constantine Manikopoulos and Symeon Papavassiliou. Network intrusion and fault detection: a statistical anomaly approach. *IEEE Communications Magazine*, 40(10):76–82, 2002.

[11] Paulo CG Costa, Kathryn B Laskey, and Ghazi Alghamdi. Bayesian ontologies in ai systems, 2006.

[12] Peng Xie, Jason H Li, Xinming Ou, Peng Liu, and Renato Levy. Using bayesian networks for cyber security analysis. In *2010 IEEE/IFIP International Conference on Dependable Systems & Networks (DSN)*, pages 211–220. IEEE, 2010.

[13] Mark S Boddy, Johnathan Gohde, Thomas Haigh, and Steven A Harp. Course of action generation for cyber security using classical planning. In *ICAPS*, pages 12–21, 2005.

[14] Namita Parati, Latesh Malik, and AG Joshi. Artificial intelligence based threat prevention and sensing engine: Architecture and design issues. In *2008 First International Conference on Emerging Trends in Engineering and Technology*, pages 304–307. IEEE, 2008.

[15] Cyrus F Nourani and RM Moudi. Intelligent cyberspace interfaces and wap generic computing. *IKS, St. Thomas, Virgin Islands*, 2002.

[16] Wei Li. Using genetic algorithm for network intrusion detection. *Proceedings of the United States department of energy cyber security group*, 1(1):8, 2004.

[17] Konrad Rieck, Philipp Trinius, Carsten Willems, and Thorsten Holz. Automatic analysis of malware behavior using machine learning. *Journal of computer security*, 19(4):639–668, 2011.

[18] Zihan Wu, Hong Zhang, Penghai Wang, and Zhibo Sun. Rtids: A robust transformer-based approach for intrusion detection system. *IEEE Access*, 10:64375–64387, 2022.

[19] H Wang, ZeZXeZBePJ Lei, X Zhang, B Zhou, and J Peng. Machine learning basics. *Deep learning*, pages 98–164, 2016.

[20] Chetan L Srinidhi, Ozan Ciga, and Anne L Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67:101813, 2021.

[21] Iqbal H Sarker, Yoosef B Abushark, Fawaz Alsolami, and Asif Irshad Khan. Intrudtree: a machine learning based cyber security intrusion detection model. *Symmetry*, 12(5):754, 2020.

[22] Nasrin Sultana, Naveen Chilamkurti, Wei Peng, and Rabei Alhadad. Survey on sdn based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications*, 12:493–501, 2019.

[23] Dieu Tien Bui, Paraskevas Tsangaratos, Viet-Tien Nguyen, Ngo Van Liem, and Phan Trong Trinh. Comparing the prediction performance of a deep learning neural network model with conventional machine learning models in landslide susceptibility assessment. *Catena*, 188:104426, 2020.

[24] Nikolaus Kriegeskorte and Tal Golan. Neural network models and deep learning. *Current Biology*, 29(7):R231–R236, 2019.

[25] Simon Du, Jason Lee, Yuandong Tian, Aarti Singh, and Barnabas Poczos. Gradient descent learns one-hidden-layer cnn: Don't be afraid of spurious local minima. In *International Conference on Machine Learning*, pages 1339–1348. PMLR, 2018.

[26] Ralf C Staudemeyer and Eric Rothstein Morris. Understanding lstm–a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*, 2019.

[27] Mahmoud Said Elsayed, Nhien-An Le-Khac, Soumyabrata Dev, and Anca Delia Jurcut. Network anomaly detection using lstm based autoencoder. In *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pages 37–45, 2020.

[28] Kai Han, An Xiao, Enhua Wu, Jianyuan Guo, Chunjing Xu, and Yunhe Wang. Transformer in transformer. *Advances in Neural Information Processing Systems*, 34:15908–15919, 2021.

[29] Tuan Anh Tang, Lotfi Mhamdi, Des McLernon, Syed Ali Raza Zaidi, Mounir Ghogho, and Fadi El Moussa. Deepids: Deep learning approach for intrusion detection in software defined networking. *Electronics*, 9(9):1533, 2020.

[30] Abebe Abeshu Diro and Naveen Chilamkurti. Distributed attack detection scheme using deep learning approach for internet of things. *Future Generation Computer Systems*, 82:761–768, 2018.

[31] Chuanlong Yin, Yuefei Zhu, Jinlong Fei, and Xinzheng He. A deep learning approach for intrusion detection using recurrent neural networks. *Ieee Access*, 5:21954–21961, 2017.

[32] Jihyun Kim, Jaehyun Kim, Huong Le Thi Thu, and Howon Kim. Long short term memory recurrent neural network classifier for intrusion detection. In *2016 international conference on platform technology and service (PlatCon)*, pages 1–5. IEEE, 2016.

[33] Partha S. Sarker, Md. Fazley Rafy, Anurag K. Srivastava, and R. K. Singh. Cyber anomaly-aware distributed voltage control with active power curtailment and ders. *IEEE Transactions on Industry Applications*, pages 1–12, 2023.

[34] KP Tripathi. A review on knowledge-based expert system: concept and architecture. *IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches & Practical Applications*, 4:19–23, 2011.

[35] Mohamed Amine Ferrag, Leandros Maglaras, Ahmed Ahmim, Makhlouf Derdour, and Helge Janicke. Rdtids: Rules and decision tree-based intrusion detection system for internet-of-things networks. *Future internet*, 12(3):44, 2020.

[36] Zakiyabanu S Malek, Bhushan Trivedi, and Axita Shah. User behavior pattern-signature based intrusion detection. In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pages 549–552. IEEE, 2020.

[37] Joseph Bamidele Awotunde, Chinmay Chakraborty, and Abidemi Emmanuel Adeniyi. Intrusion detection in industrial internet of things network-based on deep learning model with rule-based feature selection. *Wireless communications and mobile computing*, 2021:1–17, 2021.

[38] Iqbal H Sarker, Md Hasan Furhad, and Raza Nowrozy. Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science*, 2:1–18, 2021.

[39] Said Salloum, Tarek Gaber, Sunil Vadera, and Khaled Shaalan. Phishing email detection using natural language processing techniques: a literature survey. *Procedia Computer Science*, 189:19–28, 2021.

[40] Tiberiu-Marian Georgescu. Natural language processing model for automatic analysis of cybersecurity-related documents. *Symmetry*, 12(3):354, 2020.

[41] Otgonpurev Mendsaikhan, Hirokazu Hasegawa, Yukiko Yamaguchi, and Hajime Shimada. Identification of cybersecurity specific content using the doc2vec language model. In *2019 IEEE 43rd annual computer software and applications conference (COMPSAC)*, volume 1, pages 396–401. IEEE, 2019.

[42] Stephen Adams, Bryan Carter, Cody Fleming, and Peter A Beling. Selecting system specific cybersecurity attack patterns using topic modeling. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 490–497. IEEE, 2018.

[43] Taneeya Satyapanich, Francis Ferraro, and Tim Finin. Casie: Extracting cybersecurity event information from text. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8749–8757, 2020.

[44] Ram Shankar Siva Kumar, Magnus Nyström, John Lambert, Andrew Marshall, Mario Goertzel, Andi Comissoneru, Matt Swann, and Sharon Xia. Adversarial machine learning-industry perspectives. In *2020 IEEE Security and Privacy Workshops (SPW)*, pages 69–75. IEEE, 2020.

[45] Katina Michael, Roba Abbas, Payyazhi Jayashree, Ruwan J Bandara, and Anas Aloudat. Biometrics and ai bias. *IEEE Transactions on Technology and Society*, 3(1):2–8, 2022.

[46] Andrew McStay. Emotional ai, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, 7(1):2053951720904386, 2020.

[47] Jacob Sakhnini and Hadis Karimipour. Ai and security of cyber physical systems: Opportunities and challenges. *Security of Cyber-Physical Systems: Vulnerability and Impact*, pages 1–4, 2020.

[48] Vasileios Koutsouvelis, Stavros Shiaeles, Bogdan Ghita, and Gueltoum Bendiab. Detection of insider threats using artificial intelligence and visualisation. In *2020 6th IEEE Conference on Network Softwarization (NetSoft)*, pages 437–443. IEEE, 2020.

[49] Marzieh Bitaab and Sattar Hashemi. Hybrid intrusion detection: Combining decision tree and gaussian mixture model. In *2017 14th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC)*, pages 8–12. IEEE, 2017.

[50] Xinyu Yang, Peng Zhao, Xialei Zhang, Jie Lin, and Wei Yu. Toward a gaussian-mixture model-based detection scheme against data integrity attacks in the smart grid. *IEEE Internet of Things Journal*, 4(1):147–161, 2016.

[51] Chia-Mei Chen, Dah-Jyh Guan, Yu-Zhi Huang, and Ya-Hui Ou. Anomaly network intrusion detection using hidden markov model. *Int. J. Innov. Comput. Inform. Control*, 12:569–580, 2016.

[52] Amir Sinaeepourfard, Souvik Sengupta, John Krogstie, and Ricardo Ruiz Delgado. Cybersecurity in large-scale smart cities: novel proposals for anomaly detection from edge to cloud. In *2019 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*, pages 130–135. IEEE, 2019.

[53] Md Amran Siddiqui, Jack W Stokes, Christian Seifert, Evan Argyle, Robert McCann, Joshua Neil, and Justin Carroll. Detecting cyber attacks using anomaly detection with explanations and expert feedback. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2872–2876. IEEE, 2019.

[54] Li Yang, Abdallah Moubayed, Abdallah Shami, Parisa Heidari, Amine Boukhtouta, Adel Larabi, Richard Brunner, Stere Preda, and Daniel Migault. Multi-perspective content delivery networks security framework using optimized unsupervised anomaly detection. *IEEE Transactions on Network and Service Management*, 19(1):686–705, 2021.

[55] Abdulrahman Takiddin, Muhammad Ismail, Usman Zafar, and Erchin Serpedin. Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids. *IEEE Systems Journal*, 16(3):4106–4117, 2022.

[56] Partha S Sarker, Md Fazley Rafy, Anurag K Srivastava, and RK Singh. Cyber anomaly-aware distributed voltage control with active power curtailment and ders. *IEEE Transactions on Industry Applications*, 2023.

[57] Matthew Baker, Amin Y Fard, Hassan Althuwaini, and Mohammad B Shadmand. Real-time ai-based anomaly detection and classification in power electronics dominated grids. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 4(2):549–559, 2022.

[58] Priti Prabhakar, Sujata Arora, Anita Khosla, Rajender Kumar Beniwal, Moses Ndole Arthur, José Luis Arias-Gonzáles, Franklin Ore Areche, et al. Cyber security of smart metering infrastructure using median absolute deviation methodology. *Security and Communication Networks*, 2022, 2022.

[59] Manish Choubisa, Ruchi Doshi, Narendra Khatri, and Kamal Kant Hiran. A simple and robust approach of random forest for intrusion detection system in cyber security. In *2022 International Conference on IoT and Blockchain Technology (ICIBT)*, pages 1–5. IEEE, 2022.

[60] E.A. Giakoumakis, G. Papaconstantinou, and Emmanuel Skordalakis. Rule-based systems and pattern recognition. *Pattern Recognition Letters*, 5:267–272, 04 1987.

[61] Abdul Waleed, Abdul Fareed Jamali, and Ammar Masood. Which open-source ids? snort, suricata or zeek. *Computer Networks*, 213:109116, 2022.

[62] Shamsu Hassan, Jin Wang, Christos Kontovas, and Musa Bashir. Modified fmea hazard identification for cross-country petroleum pipeline using fuzzy rule base and approximate reasoning. *Journal of Loss Prevention in the Process Industries*, 74:104616, 2022.

[63] Syed Rameem Zahra, Mohammad Ahsan Chishti, Asif Iqbal Baba, and Fan Wu. Detecting covid-19 chaos driven phishing/malicious url attacks by a fuzzy logic and data mining based intelligence system. *Egyptian Informatics Journal*, 23(2):197–214, 2022.

[64] Abdullah Ayub Khan, Aftab Ahmed Shaikh, Zaffar Ahmed Shaikh, Asif Ali Laghari, and Shahid Karim. Ipm-model: Ai and metaheuristic-enabled face recognition using image partial matching for multimedia forensics investigation with genetic algorithm. *Multimedia Tools and Applications*, 81(17):23533–23549, 2022.

[65] Christos L Stergiou and Kostas E Psannis. Digital twin intelligent system for industrial iot-based big data management and analysis in cloud. *Virtual Reality & Intelligent Hardware*, 4(4):279–291, 2022.

[66] Pratyusa Mukherjee, Chittaranjan Pradhan, Hrudaya Kumar Tripathy, and Tarek Gaber. Kryptoschain—a blockchain-inspired, ai-combined, dna-encrypted secure information exchange scheme. *Electronics*, 12(3):493, 2023.

[67] Manjari Singh Rathore, M Poongodi, Praneet Saurabh, Umesh Kumar Lilhore, Sami Bourouis, Wajdi Alhakami, Jude Osamor, and Mounir Hamdi. A novel trust-based security and privacy model for internet of vehicles using encryption and steganography. *Computers and Electrical Engineering*, 102:108205, 2022.

[68] Ahmed Elhadad, Fulayjan Alanazi, Ahmed I Taloba, Amr Abozeid, et al. Fog computing service in the healthcare monitoring system for managing the real-time notification. *Journal of Healthcare Engineering*, 2022, 2022.

[69] Pankaj Chandre, Parikshit Mahalle, and Gitanjali Shinde. Intrusion prevention system using convolutional neural network for wireless sensor network. *Int J Artif Intell ISSN*, 2252(8938):8938, 2022.

[70] Meraj Farheen Ansari, Pawan Kumar Sharma, and Bibhu Dash. Prevention of phishing attacks using ai-based cybersecurity awareness training. *Prevention*, 2022.

[71] Andrzej MJ Skulimowski and Paweł Łydek. Adaptive design of a cyber-physical system for industrial risk management decision support. In *2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 90–97. IEEE, 2022.

[72] VS Devi Priya and S Sibi Chakkaravarthy. Containerized cloud-based honeypot deception for tracking attackers. *Scientific Reports*, 13(1):1437, 2023.

[73] Zhenyuan Li, Jun Zeng, Yan Chen, and Zhenkai Liang. Attackg: Constructing technique knowledge graph from cyber threat intelligence reports. In *European Symposium on Research in Computer Security*, pages 589–609. Springer, 2022.

[74] Sk Tanzir Mehedi, Adnan Anwar, Ziaur Rahman, Kawsar Ahmed, and Rafiqul Islam. Dependable intrusion detection system for iot: A deep transfer learning based approach. *IEEE Transactions on Industrial Informatics*, 19(1):1006–1017, 2022.

[75] Moataz Abdelkhalek, Gelli Ravikumar, and Manimaran Govindarasu. Ml-based anomaly detection system for der communication in smart grid. In *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2022.

[76] Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *European Conference on Computer Vision*, pages 303–319. Springer, 2022.

[77] Xueying Zhan, Qingzhong Wang, Kuan-hao Huang, Haoyi Xiong, Dejing Dou, and Antoni B Chan. A comparative survey of deep active learning. *arXiv preprint arXiv:2203.13450*, 2022.

[78] R Sri Skandha Moorthy and N Nathiya. Botnet detection using artificial intelligence. *Procedia Computer Science*, 218:1405–1413, 2023.

[79] Mercy Ejura Dapel, Mary Asante, Chijioke Dike Uba, and Michael Opoku Agyeman. Artificial intelligence techniques in cybersecurity management. In *Cybersecurity in the Age of Smart Societies: Proceedings of the 14th International Conference on Global Security, Safety and Sustainability, London, September 2022*, pages 241–255. Springer, 2023.

[80] Hassan A Alterazi, Pravin R Kshirsagar, Hariprasath Manoharan, Shitharth Selvarajan, Nawaf Alhebaishi, Gautam Srivastava, and Jerry Chun-Wei Lin. Prevention of cyber security with the internet of things using particle swarm optimization. *Sensors*, 22(16):6117, 2022.

[81] Gokul Yenduri, Gautam Srivastava, Praveen Kumar Reddy Maddikunta, Rutvij H Jhaveri, Weizheng Wang, Athanasios V Vasilakos, Thippa Reddy Gadekallu, et al. Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *arXiv preprint arXiv:2305.10435*, 2023.

[82] S Shreyashree, Pramod Sunagar, S Rajarajeswari, and Anita Kanavalli. A literature review on bidirectional encoder representations from transformers. *Inventive Computation and Information Technologies: Proceedings of ICICIT 2021*, pages 305–320, 2022.

[83] Ömer Aslan, Semih Serkant Aktuğ, Merve Ozkan-Okay, Abdullah Asim Yilmaz, and Erdal Akin. A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. *Electronics*, 12(6):1333, 2023.

[84] Ankit Thakkar and Ritika Lohiya. A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions. *Artificial Intelligence Review*, 55(1):453–563, 2022.

[85] Eman J Khaleefa and Dhahair A Abdulah. Concept and difficulties of advanced persistent threats (apt): Survey. *International Journal of Nonlinear Analysis and Applications*, 13(1):4037–4052, 2022.

[86] Rajesh Kumar, Rohan Kela, Siddhant Singh, and Rolando Trujillo-Rasua. Apt attacks on industrial control systems: A tale of three incidents. *International Journal of Critical Infrastructure Protection*, 37:100521, 2022.

[87] Daniel E Capano. Throwback attack: How notpetya accidentally took down global shipping giant maersk: Do you debate risks vs. cost of cybersecurity technologies, processes and training? maersk estimated notpetya costs at $250-300 million. *Control Engineering*, 70(4):39–42, 2023.

[88] Tianqing Zhu, Dayong Ye, Zishuo Cheng, Wanlei Zhou, and S Yu Philip. Learning games for defending advanced persistent threats in cyber systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(4):2410–2422, 2022.

[89] Shasha Yu and Fiona Carroll. Implications of ai in national security: Understanding the security issues and ethical challenges. In *Artificial Intelligence in Cyber Security: Impact and Implications: Security Challenges, Technical and Ethical Issues, Forensic Investigative Challenges*, pages 157–175. Springer, 2022.

[90] Doowon Jeong. Artificial intelligence security threat, crime, and forensics: Taxonomy and open issues. *IEEE Access*, 8:184560–184574, 2020.

[91] Jingfeng Zhang, Xilie Xu, Bo Han, Gang Niu, Lizhen Cui, Masashi Sugiyama, and Mohan Kankanhalli. Attacks which do not kill training make adversarial learning stronger. In *International conference on machine learning*, pages 11278–11287. PMLR, 2020.

[92] Ishai Rosenberg, Asaf Shabtai, Yuval Elovici, and Lior Rokach. Adversarial machine learning attacks and defense methods in the cyber security domain. *ACM Computing Surveys (CSUR)*, 54(5):1–36, 2021.

[93] Han Qiu, Tian Dong, Tianwei Zhang, Jialiang Lu, Gerard Memmi, and Meikang Qiu. Adversarial attacks against network intrusion detection in iot systems. *IEEE Internet of Things Journal*, 8(13):10327–10335, 2020.

[94] Briland Hitaj, Giuseppe Ateniese, and Fernando Perez-Cruz. Deep models under the gan: information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, pages 603–618, 2017.

[95] Abdul Basit, Maham Zafar, Xuan Liu, Abdul Rehman Javed, Zunera Jalil, and Kashif Kifayat. A comprehensive survey of ai-enabled phishing attacks detection techniques. *Telecommunication Systems*, 76:139–154, 2021.

[96] Shehzeen Hussain, Paarth Neekhara, Malhar Jere, Farinaz Koushanfar, and Julian McAuley. Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3348–3357, 2021.

[97] Ionuţ Anica-Popa, Liana Anica-Popa, Cristina Rădulescu, and Marinela Vrîncianu. The integration of artificial intelligence in retail: benefits, challenges and a dedicated conceptual framework. *Amfiteatru Economic*, 23(56):120–136, 2021.

[98] Oliver Budzinski, Victoriia Noskova, and Xijie Zhang. The brave new world of digital personal assistants: Benefits and challenges from an economic perspective. *NETNOMICS: Economic Research and Electronic Networking*, 20:177–194, 2019.

[99] Maanak Gupta, CharanKumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*, 2023.

[100] Jide S Edu, Jose M Such, and Guillermo Suarez-Tangil. Smart home personal assistants: a security and privacy review. *ACM Computing Surveys (CSUR)*, 53(6):1–36, 2020.

[101] Randall R Dipert. The ethics of cyberwarfare. In *Military Ethics and Emerging Technologies*, pages 159–185. Routledge, 2016.

[102] Jessica Dawson and Robert Thomson. The future cybersecurity workforce: Going beyond technical skills for successful cyber performance. *Frontiers in psychology*, 9:744, 2018.

[103] Feihu Xu, Xiongfeng Ma, Qiang Zhang, Hoi-Kwong Lo, and Jian-Wei Pan. Secure quantum key distribution with realistic devices. *Reviews of Modern Physics*, 92(2):025002, 2020.

[104] Konstantinos M Giannoutakis, G Spathoulas, Christos K Filelis-Papadopoulos, Anastasija Collen, Marios Anagnostopoulos, Konstantinos Votis, and Niels A Nijdam. A blockchain solution for enhancing cybersecurity defence of iot. In *2020 IEEE international conference on blockchain (blockchain)*, pages 490–495. IEEE, 2020.

[105] Ahmed A Abd El-Latif, Bassem Abd-El-Atty, Irfan Mehmood, Khan Muhammad, Salvador E Venegas-Andraca, and Jialiang Peng. Quantum-inspired blockchain-based cybersecurity: securing smart edge utilities in iot-based smart cities. *Information Processing & Management*, 58(4):102549, 2021.

[106] Nicola Capuano, Giuseppe Fenza, Vincenzo Loia, and Claudio Stanzione. Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access*, 10:93575–93600, 2022.