Received: 11 June 2022 | Revised: 26 November 2022 | Accepted: 25 December 2022 | Published on: 10 January 2023

DOI: 10.1002/spy2.295

SURVEY PAPER

WILEY

Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview

Iqbal H. Sarker[1,2,3]

[1]Department of Computer Science and Engineering, Chittagong University of Engineering & Technology, Chittagong, Bangladesh

[2]Swinburne University of Technology, Hawthorn, Victoria, Australia

[3]Security Research Institute, Edith Cowan University, Joondalup, Western Australia, Australia

Correspondence
Iqbal H. Sarker, Security Research Institute, Edith Cowan University, Joondalup, Western Australia, Australia.
Email: m.sarker@ecu.edu.au

Abstract

Due to the rising dependency on digital technology, cybersecurity has emerged as a more prominent field of research and application that typically focuses on securing devices, networks, systems, data and other resources from various cyber-attacks, threats, risks, damages, or unauthorized access. Artificial intelligence (AI), also referred to as a crucial technology of the current Fourth Industrial Revolution (Industry 4.0 or 4IR), could be the key to *intelligently* dealing with these cyber issues. Various forms of AI methodologies, such as analytical, functional, interactive, textual as well as visual AI can be employed to get the desired cyber solutions according to their computational capabilities. However, the dynamic nature and complexity of real-world situations and data gathered from various cyber sources make it challenging nowadays to build an effective AI-based security model. Moreover, defending *robustly* against adversarial attacks is still an open question in the area. In this article, we provide a comprehensive view on "Cybersecurity Intelligence and Robustness," emphasizing multi-aspects *AI-based modeling* and *adversarial learning* that could lead to addressing diverse issues in various cyber *applications* areas such as detecting malware or intrusions, zero-day attacks, phishing, data breach, cyberbullying and other cybercrimes. Thus the eventual security modeling process could be automated, intelligent, and robust compared to traditional security systems. We also emphasize and draw attention to the future aspects of cybersecurity intelligence and robustness along with the *research direction* within the context of our study. Overall, our goal is not only to explore AI-based modeling and pertinent methodologies but also to focus on the resulting model's applicability for securing our digital systems and society.

KEYWORDS

adversarial attacks, artificial intelligence, cyber data analytics, cybersecurity, Industry 4.0 applications, intelligent decision-making, machine learning, robust secured systems

1 | INTRODUCTION

Technology is now more important than it has ever been in the modern world. Unauthorized access,[1] malware attacks,[2] zero-day attacks,[3] data breach,[4] denial of service (DoS),[1] social engineering or phishing,[5] cyberbullying[6] and other security incidents, discussed briefly in Section 2 have increased at an exponential rate in recent years due to the increasing reliance on digitization and Internet-of-Things (IoT).[7] An intentional attempt to compromise the information system of another individual or organization is typically known as a cyber-attack. These security incidents also referred to as cybercrime, can cause harm to a company and its personnel as well as cause interruption and financial losses. For instance, a data breach in the United States costs 8.19 million USD on average, according to IBM security reports,[8] while cybercrime damages the world economy by 400 billion USD annually.[9] Similarly, according to the statistical information of Juniper Research,[10] the number of records breached each year is predicted to roughly triple over the next 5 years. In order to minimize the loss, businesses and organizations need to think and put into action a strong cybersecurity strategy. Cybersecurity experts and researchers nowadays are being warned by the exponential growth of such types of cybercrimes mentioned above. Therefore, *protecting* an information system, especially one that is connected to the Internet, from various cyber-attacks, threats, risks, damages, or unauthorized access is a crucial issue that needs to be handled intelligently and automatically, in which we are interested.

The major downside of today's digital era is the cybersecurity risk as security breaches are becoming more frequent and destructive. In the real-world scenario, a nation's entire national security typically depends on a security management system that can identify and prevent security issues efficiently and intelligently. Today's organizations in the government, national guard, commercial, financial, medical, and other sectors frequently collect, process, and store enormous amounts of data on computers, servers and other devices, demanding the need for smart cybersecurity services and management.[11,12] According to the definition of cybersecurity, it is a collection of technologies, methods, and practices for protecting devices, networks, programs, and data from attacks, damages, and unauthorized access; also known as "information technology security" or "electronic information security."[11] Antivirus, firewalls, user authentication, encryption, and other well-known security solutions may not be sufficient to meet today's diverse needs.[13-16] The underlying issue with traditional systems is that they are frequently maintained by a small number of knowledgeable security professionals, and data processing is done on an ad-hoc basis, limiting them from responding intelligently to needs.[17,18] To solve this problem, we need to develop more responsive and effective security systems that can automatically react to cyber-attacks and intelligently mitigate them through strong security policies in real-time, where artificial intelligence (AI) can play a key role from the technical point of view.

As one of the core technologies of today's Fourth Industrial Revolution (Industry 4.0 or 4IR), AI has the processing power and capability to be a key component of better cybersecurity services through intelligent decision-making and automation. AI is typically known as a vast field of computer science that is concerned with building intelligent machines that can perform tasks that normally require human intelligence.[19] In other words, it aims to make computers smarter and more intelligent by enabling them to think and learn via computer programs or machines, allowing them to think and function similarly to humans. Therefore, the primary objective of AI is to make it possible for computers and other devices to carry out cognitive tasks including problem-solving, decision-making, perception, and communication interpretation. However, building an effective AI model is a challenging endeavor because of the dynamic nature and complexity of real-world cyber problems and data. To find an effective AI-based solution, it is necessary to comprehend the nature of the issue and carry out thorough research to solve a real-world issue in the context of cybersecurity, for example, identifying cyber-anomalies or multi-attacks.[20] In addition, how to effectively protect against adversarial attacks, when attackers intend to underperform the model, is still an open question. Popular AI techniques including machine learning (ML) methods, for instance, are used to intelligently tackle today's numerous cybersecurity challenges.[21,22] In this article, we provide multi-aspects AI-based security modeling and adversarial ML that could lead to addressing diverse issues in various cyber applications as well the future aspect in terms of automation, intelligent decision-making, and robustness in the context of today's cybersecurity systems and society.

To achieve this goal, various forms of AI, such as analytical, functional, interactive, textual as well as visual AI are being explored in this article. These can be employed depending on the nature of the target cyber problem and the intended solutions. Although AI is a large topic, we concentrate on potential AI-based solutions according to their computing capabilities for solving real-world cybersecurity issues, with the outcome being used to build automated and intelligent systems in a variety of application areas. To build AI-based security modeling by taking into account today's cybersecurity needs, we classify the above-mentioned AI strategies into 10 categories: "(i) ML; (ii) neural networks and deep learning

(DL); (iii) data mining, knowledge discovery, and advanced analytics; (iv) rule-based modeling and decision-making; (v) fuzzy logic-based approach; (vi) knowledge representation, uncertainty reasoning, and expert system modeling; (vii) case-based reasoning (CBR); (viii) text mining and natural language processing (NLP); (ix) visual analytics, computer vision and pattern recognition, and (x) hybridization, searching and optimization," discussed briefly in Section 3. These methods can be incredibly useful in intelligently resolving cybersecurity issues including malware detection, zero-day attacks, anomaly or fraud detection, and other cybercrimes, depending on the nature of the problem and the intended solution. In addition, adversarial ML typically emphasizes how ML algorithms are attacked and how they are defended, discussed briefly in Section 4. Thus, it's crucial to comprehend the principles underlying these methods as well as how these can be used in various real-world application scenarios, summarized in Section 5. Therefore, the purpose of this article is to build the foundation for academics and industry people who seek to study, investigate, and develop automated and intelligent systems in the field of cybersecurity utilizing multi-aspects AI methodologies.

The main contributions of this article are therefore listed as follows:

- To motivate cybersecurity research and development towards intelligence and robustness in light of current real-world requirements and define the scope of our study accordingly.
- To explore multi-aspects of AI-based modeling such as analytical, functional, interactive, textual, and visual AI, as well as adversarial learning in order to comprehend the theme of the effective use of AI methods towards cybersecurity automation with intelligent decision-making, and robustness in security modeling.
- To explore the applicability of AI-based security modeling in various real-world scenarios and issues in order to aid researchers and professionals in broadening their perspectives on AI methodologies in the context of cybersecurity.
- To identify and outline the future aspect of cybersecurity along with research directions within the scope of our study in order to undertake future research and development.

The rest of the article is carried out in the following order. Section 2 explains why AI is being researched and used today in the context of cybersecurity. We explore and outline how different AI techniques can be employed for security intelligence modeling in various cybersecurity concerns in Section 3. We also explore adversarial ML in Section 4. We highlight different real-world application domains where AI techniques can be used for cybersecurity intelligence and robustness in Section 5 as well as the future aspect of cybersecurity highlighting research issues, and finally, Section 6 brings this article to an end.

2 | UNDERSTANDING CYBERSECURITY INTELLIGENCE AND ROBUSTNESS

In this section, we define cybersecurity intelligence and robustness, which is based on AI methods as well as explore its background and related components within the scope of our study.

2.1 | Defining cybersecurity

During the recent half-century, our modern and digital civilization has been more interconnected with information and communication technologies (ICT). The prevalence of data breaches and attacks is growing as a result of the fact that the majority of the smart computers we use daily are powered by global Internet access. Therefore, ICT security—the detection and defense of ICT systems against various types of advanced cyber-attacks or threats has been a top priority for our security professionals or policymakers in recent years.[23] Enterprises use ICT security to ensure the integrity, confidentiality, and availability of their data and systems by implementing safeguards, policies, and processes. Simply said, cybersecurity is the process of protecting things that are vulnerable due to the use of ICT. Cybersecurity is a term that has a variety of different meanings and is widely used nowadays; several key terms such as "information security," "data security," "network security," and "Internet or IoT security"[24] are frequently interchanged, confusing readers and professionals in the field. Among these, the term "cybersecurity" has higher global popularity than others and is growing day by day.[21]

Cybersecurity has been characterized in a variety of ways by various researchers. For example, cybersecurity refers to the various activities or policies that are implemented to protect ICT systems from threats or attacks.[9] Craigen et al[25] defined "cybersecurity as a set of tools, practices, and guidelines that can be used to protect computer networks, software programs, and data from attack, damage, or unauthorized access." According to Aftergood et al,[26] "cybersecurity is a set of technologies and processes designed to protect computers, networks, programs and data from attacks and unauthorized access, alteration, or destruction." Overall, cybersecurity is concerned with identifying various cyber-attacks or threats as well as the related defense tactics to prevent them and ultimately, secure the systems, which is associated with confidentiality, integrity, and availability.[27-29] The CIA triad[21] exploring confidentiality, integrity, and availability as mentioned, is the core principles used to drive information security policy within an enterprise, where the individual losses of these principles or their combinations are considered a threat. Such cyber threats are also known as "Data theft," "Data alteration," and "Denial access of data" respectively as shown in Figure 1. Overall, based on the CIA triad for the security policy stated above, we can conclude that "Confidentiality" protects data, objects, and resources from unauthorized access and misuse; "Integrity" protects data from unauthorized changes, and "Availability" ensures accessibility to the systems and the resources to authorized users or the appropriate entity.

2.2 | Examples of security incident and attacks

A security incident is usually a malicious act that directly compromises the security aspects of CIA triad principles described before, can cause harm to a company or an individual.[1] In the cyber context, a threat is generally defined as a probable security breach that could take control of a system or asset's vulnerability, whereas an attack is defined as an intentional, unauthorized action against a system or asset. Malware, phishing, data breaches, ransomware, cyberbullying, social engineering, denial of service (DoS), zero-day attacks and so forth are some examples of cyber-attacks. The most prominent cyber threats and attacks that need to be considered in today's cyber environment are listed below.

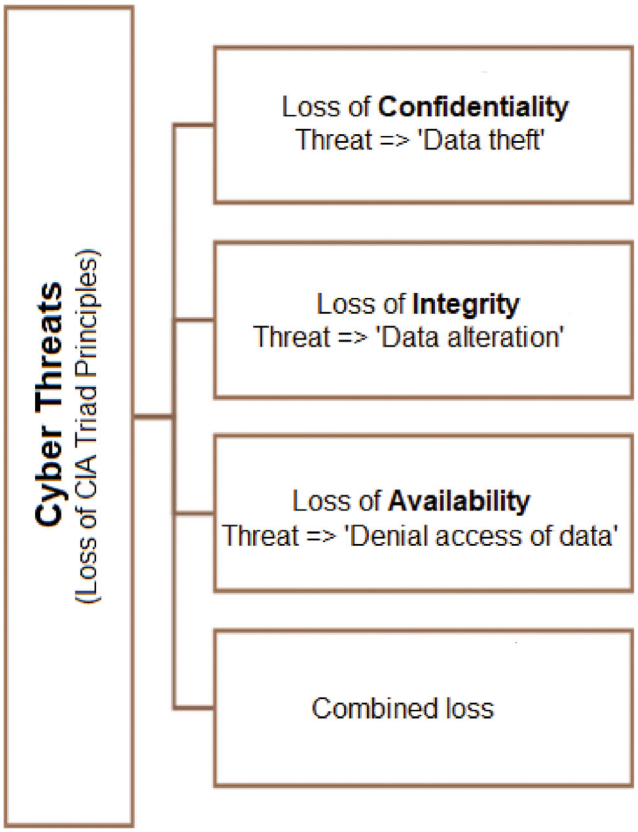

FIGURE 1 An illustration of cyber threats with the loss of CIA (confidentiality, integrity, and availability) triad principles used to drive information security policy within an enterprise

- *Cyber risk, threat, and vulnerability*: In cybersecurity, risk refers to the possibility of losing, damaging, or destroying assets or data. A negative event, such as the exploit of vulnerability, is referred to as a threat. Vulnerability, on the other hand, is a weakness that exposes one to threats and so raises the possibility of a negative event. Thus, in terms of the relationship among these terms, we can say that when a threat exploits a vulnerability in an IT infrastructure, network, or application, it puts the assets, data, and business at risk.[11]
- *Cyber-attack*: Any type of offensive activity that tries to steal, expose, alter, disable, or destroy information through gaining unauthorized access to computer systems is referred to as a cyber-attack. Malware, phishing, denial-of-service (DoS) attacks, SQL injection, zero-day exploits, and other threats are some of the examples.[11]
- *Malware*: Malware, short for "malicious software," is an intrusive program created by cyber criminals, commonly referred to as "hackers," to steal data and damage or destroy computers and computer systems. Malware is one of the most serious security threats on the Internet these days as most online threats are some form of malware. Viruses, worms, trojans, spyware, adware, ransomware, backdoor, malicious bot and so forth are examples of widespread malware.[28,30,31]
- *Ransomware*: Ransomware is a sort of malware attack in which the attacker encrypts and locks the data and important files of the victim, then demands payment to unlock and decrypt the data.[2]
- *Data breach*: A data breach is a security violation that occurs when an individual, group, or software system gains unauthorized access to and retrieves confidential, sensitive, or protected information, such as financial data.[11]
- *Intrusion*: Any unauthorized activity on a digital network is referred to as a network intrusion. Thus a system that can monitor and react to a network or systems for such malicious activities or policy violations is known as an intrusion detection system.[32]
- *Unauthorized access*: Any access that violates the established security policy by the owner or operator of a computer system or network is considered unauthorized access.[1]
- *Anomaly*: Any deviations from the typical data points, items, observations, or events that do not reflect the expected pattern are known as anomalies. Although these anomalies are unlikely, they might be signs of a substantial risk like fraud or cyber breaches.[20]
- *Social engineering and phishing*: Social engineering is a broad term that describes many malevolent acts carried out through human interactions that rely on psychological manipulation to convince users to disclose sensitive information or commit security errors. Phishing is a technique of social engineering in which malicious websites or emails are typically used to obtain sensitive information, frequently financial information.[28,33]
- *Cyberbullying*: The use of technology to harass, threaten, embarrass, or target another person is known as cyberbullying. It can happen on social media, messaging apps, gaming apps, and mobile phones.[6]
- *Zero-day attack*: The term "zero days" or "day zero attack" refers to the threat of an unknown security vulnerability in computer systems or an application for which a patch has not been provided or for which the application developers were unaware of or did not have enough time to resolve.[34]
- *Denial-of-service (DoS)*: A cyberattack known as a "denial-of-service attack" aims to make a computer or network resource unavailable to the users who are expected to utilize it by temporarily or permanently disrupting the services of a host connected to a network.[28]
- *Cryptocurrency issues and blockchain*: A cryptocurrency is a digital, encrypted, and decentralized medium of exchange that is based on blockchain technology, such as Bitcoin, Litecoin, Ethereum, Cardano, and others. phishing and social engineering attacks are the most popular forms of crypto hacking. Ransomware criminals typically demand ransom in cryptocurrency, such as Bitcoin, due to its perceived secrecy and ease of online payment.[35,36]
- *Adversarial attacks*: In the context of security modeling, "adversary" typically refers to individuals or devices that could make an effort to penetrate or corrupt a computer network or program. A ML model can be disrupted by adversaries using a variety of attack methods, either before the model is trained, for example, poisoning attack, or after it has been trained, for example, evasion attack.[37-39]

Overall, cybercrime typically refers to any criminal activity involving devices and networks, such as email and Internet fraud, financial or card payment data theft, cyber harassment, privacy violations, and many others. Thus we can say that any traditional crime or relevant conducted through the Internet is a sort of cybercrime. As cybercrime becomes more sophisticated, criminals are now increasingly targeting people, businesses, organizations, educational institutions,

governments, and other crucial sectors. Thus protecting a system from various cyber-attacks, threats, risks, damages, or unauthorized access is a crucial issue that needs to be handled intelligently and automatically, where the knowledge of AI could play a significant role from the technological point of view.

2.3 | Potentiality of artificial intelligence

It is nearly hard to successfully deal with today's diverse cyber threats using only human analysis given the exponential growth in threat propagation carried on by the daily release of new malware. Thus developing algorithms that will enable automation to effectively handle such issues. The use of AI techniques in a variety of cyber security applications, discussed briefly in Section 5 has recently been attempted to achieve the goal. AI is concerned with comprehending and carrying out intelligent tasks including thinking, creating new things, and adapting to new contexts and problems.[19] Various forms of AI methods including analytical, functional, interactive, textual, and visual AI, can be utilized depending on the nature of the problem and the desired consequences as below.

- *"Analytical AI"*—is usually focused on the ability to extract insights or patterns from data to provide suggestions, assisting in data-driven decision-making.
- *"Functional AI"*—is comparable to analytical AI in that it executes an action based on extracted insights or knowledge instead of offering suggestions.
- *"Interactive AI"*—typically allows businesses to automate communication without sacrificing interaction.
- *"Textual AI"*—is concerned with text analytics and mining as well as language processing.
- *"Visual AI"*—is concerned with computer vision and extracting insights from image or visual data.
- *"Hybrid AI"*—could be concerned with the combinations of the above-mentioned AI, depending on the problem nature and target solutions in the context of cybersecurity.

In the real world, each AI type can address a different set of challenges. As briefly discussed in Section 3, various AI techniques such as ML, DL, advanced analytics, knowledge discovery, and other relevant AI techniques and their hybridization can be used to provide solutions that take into account the target applications. To provide an automated and data-driven intelligent solution that meets today's needs in the context of cybersecurity, the majority of real-world scenarios involve advanced analytics,[40] which leads to data science and analytical AI, that uses ML and DL techniques. These techniques could also be considered a frontier of AI that has the potential to develop intelligent systems and automate tasks. Figure 2 depicts the position of ML and DL in the field of AI. DL is a subset of both ML and AI, as demonstrated in Figure 2. ML[41] automates the construction of analytical models using data or experience, whereas AI[19] incorporates human behavior and intelligence into machines or systems.

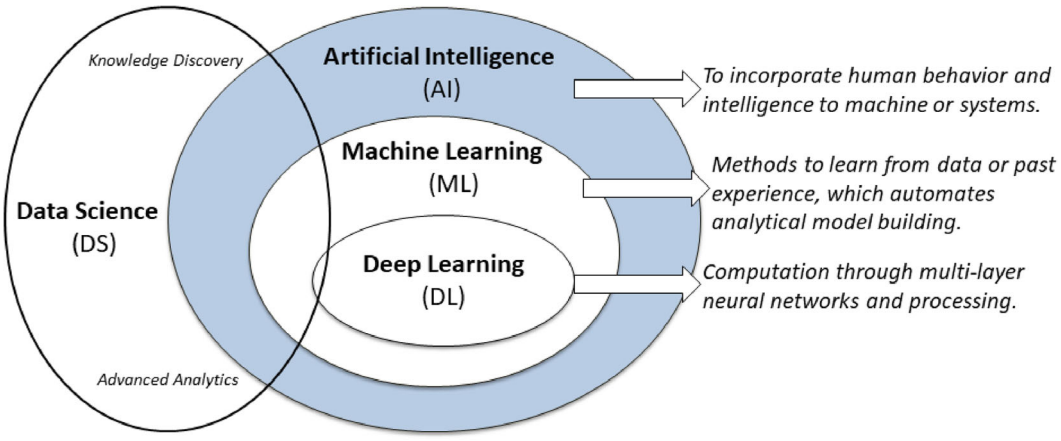

FIGURE 2 An illustration of the position of machine learning (ML) and deep learning (DL) within the area of artificial intelligence (AI) as well as with data science (DS)

Overall, these technologies have the potential to transform today's cybersecurity environment, especially in terms of a strong computing engine, as well as contribute to technology-driven automation and intelligent cybersecurity systems. Depending on the nature of the problem and the target cyber solution, several different strategies can be used to construct AI-based models to address various real-world security concerns, as addressed briefly in Section 3.

2.4 | Defining cybersecurity intelligence, automation, and robustness

The knowledge of AI is revolutionizing the cybersecurity industry around the world. Due to its computational capability for intelligent decision-making and automation in a specific problem domain, it is taken into account as the key to the development of intelligent services.[19] Traditional security solutions such as antivirus, firewalls, user authentication, encryption, and other well-known security solutions might not be sufficient to meet today's diverse needs.[13-16] As mentioned earlier, the main issue with traditional systems is that they are typically maintained by a small number of knowledgeable security professionals, and the data processing is done on an ad-hoc basis, limiting them from responding intelligently and automatically to today's needs.[17,18] These two properties "automation" and "intelligence" as well as "robustness" in the context of cybersecurity is defined as below:

- *Automation in cybersecurity*: Cybersecurity automation typically reduces human dependency while providing a better and more methodical means to detect, protect against, respond to, and recover from various security attacks. Cybersecurity can be automated with the use of cutting-edge technologies like AI[19] and ML[41] as discussed briefly in Section 3. For example, an effective anti-malware solution is a sort of ML modeling that is used in cybersecurity defense. Thus, an automated and intelligent system can help detect and respond to cyber threats like malicious activities in real-time when connected with security management.
- *Intelligence in cybersecurity*: It's often referred to as computational intelligence, and it characterizes a system's capability to extract useful knowledge from cybersecurity data to derive conclusions or through experimental observation, as well as from learning a particular cybersecurity task. The usage of intelligent systems could help in a variety of cybersecurity issues, including malware behavior analysis, data breach prediction, zero-day attacks, fraud detection, phishing detection, and others. Therefore, as explained in Section 3 ML and data-driven modeling as well as advanced analytics are the key drivers of intelligent computing and decision-making in the context of cybersecurity. Designing new methods or variants by taking into account model optimization, accuracy, and applicability, following the type of data and the intended real-world problem could be a novel contribution to the area. Therefore, the question is, "How can we design an effective and efficient learning algorithm or model that enables learning process more precisely from patterns or features utilizing relevant cyber data?"
- *Robustness in cybersecurity*: Although the majority of research on adversarial learning has been concentrated on the computer vision domain,[42] which lacks a real adversary, the area of cyber security is especially fascinating because of the recent increase of adversaries in the cyber security space. For instance, malware developers are now looking to get beyond machine and DL-based next-generation antivirus products, spam filters and so forth. In general, the proactive security method aids in predicting cyberattacks in advance, whereas the reactive security approach focuses on minimizing the damage done by cybersecurity threats and addressing vulnerabilities, discussed briefly in Section 4. In the domain of cybersecurity, ML approaches are often employed to detect cybersecurity issues, where adversaries actively transform their objects to avoid detection. Thus defending robustly against adversarial attacks is still an open question. Therefore, the question is—"How to effectively build a robust security model to defend against diverse adversarial attacks in the cyber security space?"

Overall, multi-aspects AI and its computational strength and capabilities make it a strong fit for intelligent and robust cybersecurity services and automation, which is the foundation of the term "Cybersecurity Intelligence and Robustness," focused in this article. More specifically, the term cybersecurity "intelligence" is defined as a concept that takes into account AI principles and techniques to automatically and intelligently respond and manage diverse security incidents, and "robustness" is the term used to effectively protect against adversarial attacks when attackers intend to underperform the cyber model. To tackle these issues, in this article, we explore multi-aspects AI-based security modeling and adversarial ML that could lead to addressing diverse security issues in the context of today's cybersecurity systems and society, discussed in the following sections.

3 | AI-BASED MODELING IN CYBERSECURITY

In the area of cyber security, the number of attacks has increased rapidly than the human and financial resources that can be used to assess and combat any new cyber threat. As our digital world gets more widespread, there is a lot of sensitive, financial, and personal data that has to be protected against cyberattacks. In this section, we briefly discuss the principles and capabilities of potential AI approaches that can play a significant role in cybersecurity modeling to address a range of real-world issues. We divide AI strategies into 10 potential categories, taking into account the various types of AI mentioned earlier. These categories of AI techniques outlined below could be useful in automating and intelligently managing cybersecurity systems, depending on the nature of real-world cyber problems.

3.1 | Machine learning

The use of ML, which allows computers to be programmed to recognize patterns in data and make predictions that are more accurate over time, has increased significantly in recent years. AI relies on ML as a fundamental enabling technology, which is used for useful applications like malware or fraud detection, email spam filters, risk prediction and so forth. ML models are typically composed of a set of rules, procedures, or sophisticated "transfer functions" that may be used to uncover interesting data patterns or make predictions.[43] ML, also referred to as predictive analytics, is a sort of analytics that employs data to make predictions about yet-unknown future variables. It is used to address a wide range of practical business issues, for example, the prediction of company risk.[41] In Figure 3, a ML-based prediction model's overall structure is shown. In phase 1, the model is trained using historical data, and in phase 2, the result is generated using new test data. Depending on their learning principles and capabilities, as described below, several types of ML algorithms can be used for modeling in a particular issue area.

- *Supervised learning*: This is done when certain goals are specified to be achieved from a collection of inputs, such as when a "task-driven strategy" is used to train algorithms to classify data or forecast outcomes, such as recognizing spam-like emails. Classification, that is, predicting a label, and regression, that is, predicting a quantity, analyses are the two most typical supervised learning problems, which we briefly discussed in our previous study Sarker et al.[41] Navies Bayes,[44] KNN,[45] SVM,[46] Decision Trees—ID3,[47] C4.5,[48] CART,[49] BehavDT,[50] IntrudTree,[32] ensemble learning, random forest,[51] linear regression,[52] support vector regression[46] and so forth are some of the most often used techniques for solving various supervised learning tasks, depending on the nature of the data provided in a given problem domain.[41] For example, classification models could be beneficial in detecting different forms of cyber-attacks, whereas a regression model could be useful in analyzing cyber-crime trends or estimating financial damage in the domain of cybersecurity,

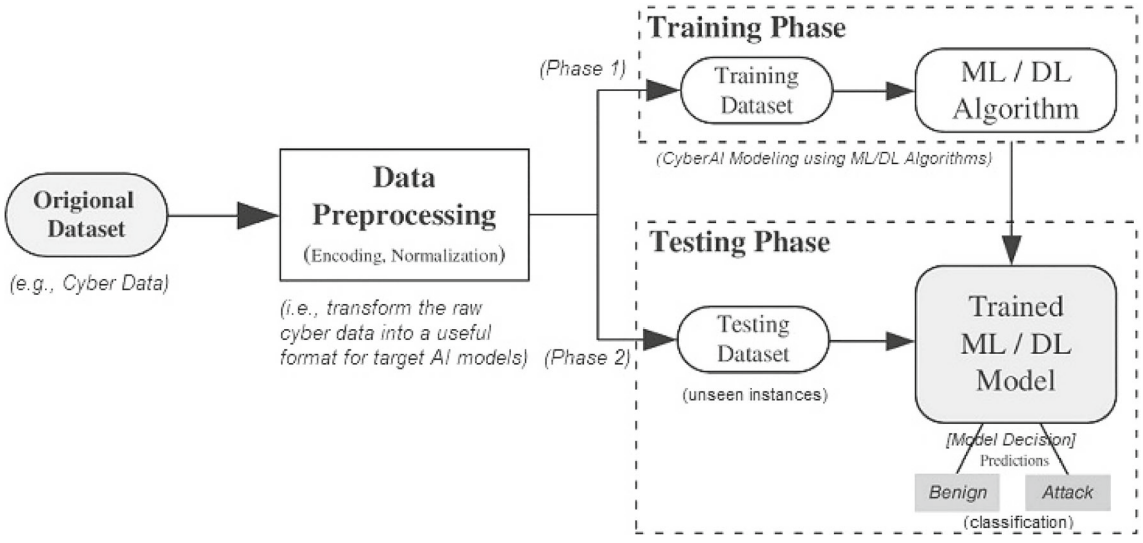

FIGURE 3 A general structure of a machine learning-based predictive model considering both the training and testing phase

allowing businesses to analyze and manage their cyber risk. A spam filter can be considered an example of a classification task to distinguish spam from other messages. When examples for specific groups are known, supervised learning is useful for classification. Through regression analysis, characteristics such as the total amount of suspicious transactions, location and so forth are used to estimate the likelihood of fraudulent activity.

- *Unsupervised learning*: In the real world, security data is not always labeled or categorized. Unsupervised learning, or a data-driven technique, can thus be used to discover patterns, structures, or knowledge from unlabeled data.[41] Clustering, a widely used form of unsupervised learning, can uncover the datasets' hidden patterns and structures. The capability of clustering to automatically identify the classes to which the sample belongs when information about the classes is not known beforehand sets it apart from classification. The investigation of malware and forensic evidence both heavily rely on these techniques. Another interesting application for clustering is the analysis of user behavior, in which program users are grouped to see if they should belong to a specific group. Security data were categorized using clustering approaches that take into consideration certain measures of similarity in the data. Several clustering algorithms, for example, K-means,[53] K-medoids,[54] DBSCAN,[55] GMM,[56] hierarchical,[57,58] BOTS[59] and so forth can be used in such purposes. Moreover, incident response and risk management from recommendation approaches is another area where association learning techniques could be useful. Several methods such as AIS,[60] Apriori,[61] FP-Tree,[62] RARM,[63] Eclat[64] as well as the concept of recently proposed ABC-RuleMiner by Sarker et al[65] can be used for building an effective rule-based model, for example, policy-rule generation and cyber modeling.
- *Other learning techniques*: Semi-supervised learning, which utilizes both labeled and unlabeled data, can be considered a combination of the supervised and unsupervised techniques previously discussed. In terms of cybersecurity, it may be useful to boost the functionality of cybersecurity models by automatically categorizing data without human input. The reinforcement learning (RL) technique is another type of ML that typically distinguishes an agent by allowing it to create its own learning experiences through interacting directly with the environment, that is, an environment-driven approach, algorithm examples could be Q-learning, deep Q networks.[66,67] For instance, the authors[68] combine RL with a neural network classifier for detecting botnet traffic or criminal cyber activity.

Overall, various categories of ML algorithms mentioned above can be used to build an automated cybersecurity model to make predictions or decisions. The algorithms are discussed briefly in our earlier paper, Sarker et al[22] highlighting their working principles, learning capabilities, and real-world applications in the domain of cybersecurity. However, the development of secure, resilient ML systems that can withstand numerous forms of malicious attacks is a crucial and urgent endeavor. Thus adversarial ML typically explores how ML algorithms are attacked and how they are defended, discussed briefly in Section 4.

3.2 | Neural networks and deep learning

Another prominent AI technique is DL,[69] which is based on artificial neural networks (ANN) and may outperform traditional ML methods, particularly when trained on massive security datasets. Due to its layer-wise learning potential from data as shown in Figure 4, DL has recently become a hot topic in the computing world. One of the emerging DL applications is the ability to detect fraud or application compromise attempts made by malicious users at the precise moment they happen. However, the ANN model's significant time consumption due to its complexity is the fundamental problem with training it.

DL algorithms can be utilized in the cybersecurity area for a variety of objectives, including detecting network intrusions, detecting and categorizing malware traffic, backdoor assaults, and so on.[31,70,71] According to our earlier paper, Sarker et al,[69] three major categories of DL approaches as shown in Figure 5 can be employed to build DL based cybersecurity model. Discriminative deep architectures such as MLP,[56] CNN,[72] RNN,[73,74] and their variants can be used to perform various types of classification or supervised learning tasks. For example, intrusion detection,[75,76] malware detection,[77-79] phishing detection,[80] botnet detection[81] and many more have been studied in the area of discriminative deep network architectures.

On the other hand, unsupervised or generative learning is commonly used for feature learning or data generation and representation.[82,83] The GAN,[84] AE,[85] RBM,[86] SOM,[87] DBN,[88] as well as their variants, discussed briefly in our earlier paper Sarker et al,[69] can be used to solve various types of real-world issues. For example, to build threat detection model,[89] intrusion detection,[90] detecting zero-day malware[91] and so forth such types of generative models are used according to

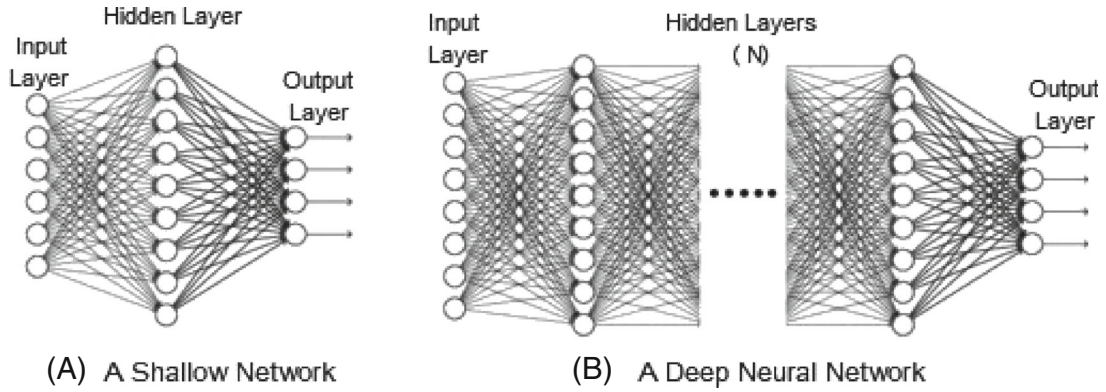

FIGURE 4 A general architecture of (A) a shallow network with one hidden layer and (B) a deep neural network with multiple hidden layers

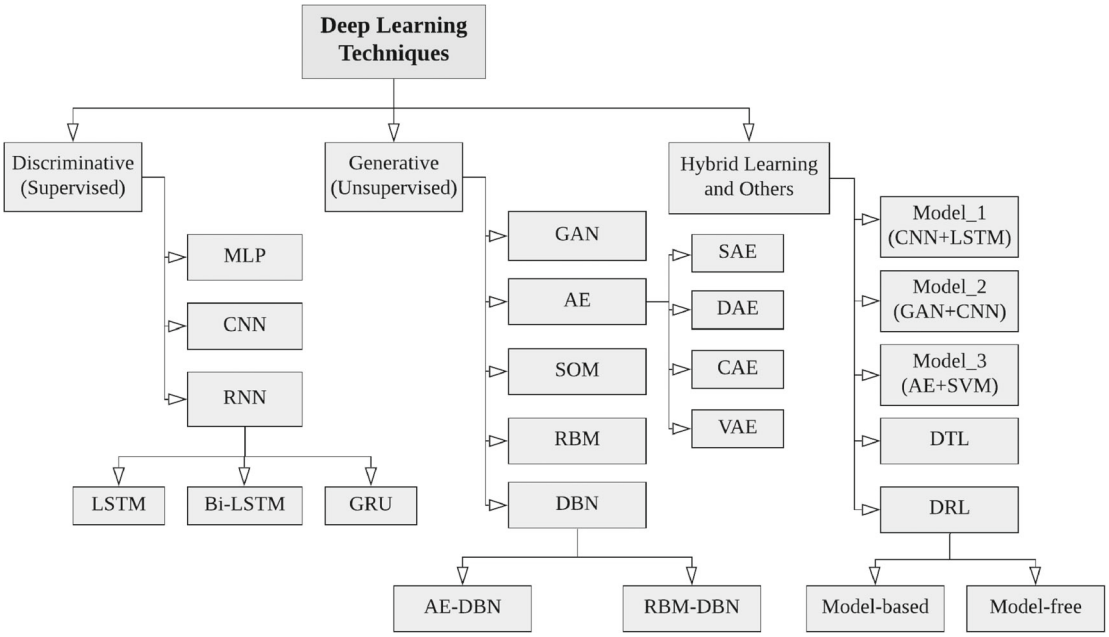

FIGURE 5 A taxonomy of DL techniques, broadly divided into three major categories (i) deep networks for supervised or discriminative learning, (ii) deep networks for unsupervised or generative learning, and (iii) deep networks for hybrid learning and relevant others

their learning capabilities. Cyber-attacks such as malware detection,[77] intrusion detection,[76,92] botnet attack[93] can also be detected using hybrid network models consisting of multiple deep basic learning models with better accuracy. In addition, a transformer network-based model can be used to solve security issues, such as autonomous cyberbullying detection,[94] intrusion or anomaly detection[95,96] and so forth.

Overall, DL models and their variations mentioned above could play a vital part in the development of effective AI models to address cybersecurity issues, depending on their learning capabilities at various levels, the nature of the data, and the desired outcome, particularly for large datasets.

3.3 | Data mining, knowledge discovery and advanced analytics

The term "data mining" has gained popularity in the last decade, with terms like "knowledge mining from data," "knowledge extraction," "knowledge discovery from data (KDD)," "data or pattern analysis" and similar others.[40] The general procedure of the knowledge discovery process, that is, security insights, is depicted in Figure 6.

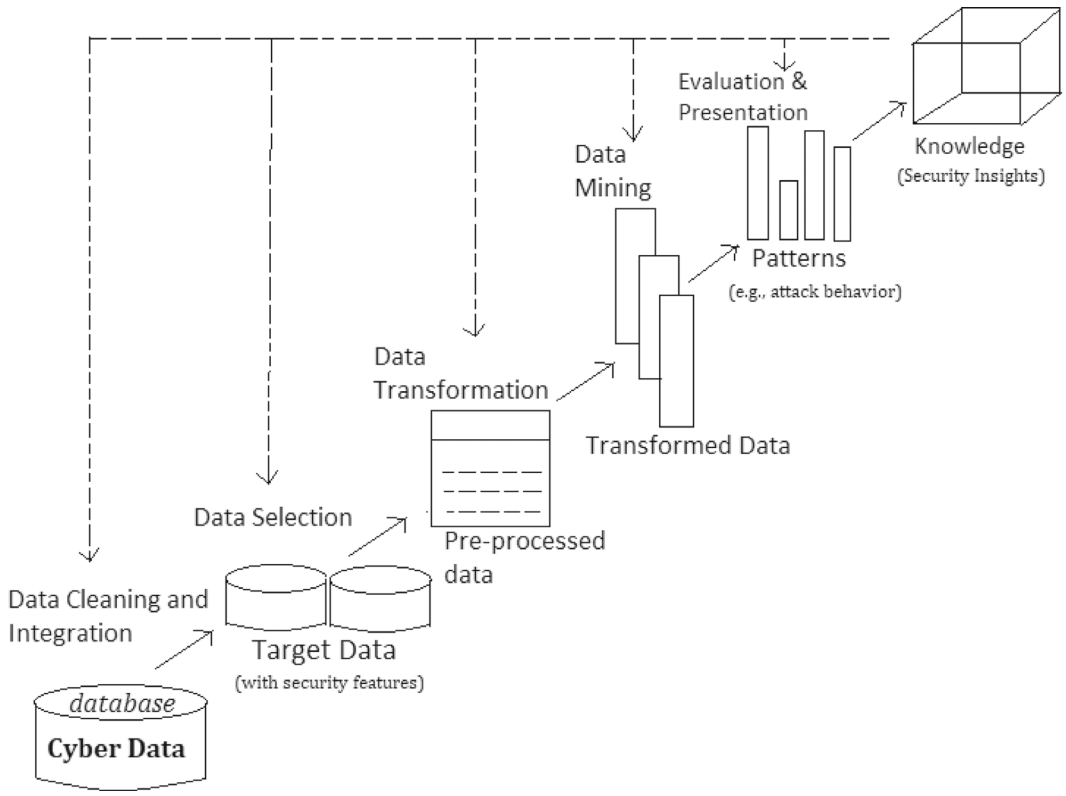

FIGURE 6 A general procedure of the knowledge discovery (security insights) process from cyber data

According to Han et al,[52] the term "knowledge mining from data" should have been used instead. Data mining, which is similar to another popular phrase "Data Science"[40] is defined as the process of extracting meaningful patterns and knowledge from large volumes of data. Data science is commonly characterized as a concept that combines statistics, data analysis, and related approaches to analyze and investigate real-world situations using data. Several key questions like "What happened?" "Why did it happen?" "What will happen in the future?" and "What action should be taken?" are all common and major queries in the field of data analytics, as discussed briefly in our earlier paper Sarker et al.[40] In the following, we have explored these questions with cybersecurity examples.

- *Descriptive analytics*: Typically answer the question "What happened in the past"? for example, describing cybersecurity risks.
- *Diagnostic analytics*: Typically answer the question "Why did it happen"? for example, diagnose vulnerabilities.
- *Predictive analytics*: Typically answer the question "What will happen in the future"? for example, predicting future malicious behavior.
- *Prescriptive analytics*: Typically answer the question "What action should be taken"? for example, prescribe protective remedies.

In summary, descriptive and diagnostic analytics look at historical data to figure out what happened and why. Analytics that are predictive and prescriptive use past data to generate predictions about what will happen in the future and what actions should be taken to reduce any negative consequences. Numerous works have been done in the area. For instance, to create a security system automated and intelligent, Sarker et al,[11] suggest a cybersecurity data science modeling that incorporates extracting security incident patterns or insights from cybersecurity data. Using data mining techniques, Tianfield et al[97] describes a strategy for detecting cyber-attacks. Li et al[98] carry out a threat intelligence-based association study of cyber-attack attribution. Thus, we can conclude that various knowledge discovery or data mining methods[40] as well as analytics could play a crucial role in the development of AI models utilizing cybersecurity data through extracting useful knowledge or security insights for a particular cyber issue.

3.4 | Rule-based modeling and decision-making

Rules are often stated as IF-THEN statements of the form:

IF < X > THEN < Y >, where X is called antecedent and Y is called the consequent. How frequently the items appear in the data is indicated by their level of support. Confidence represents how frequently the IF-THEN assertions are verified as true. How often an IF-THEN statement is anticipated to be true can be determined using a third metric called lift that compares confidence to predicted confidence.[61,99]

The term "rule-based system" was previously used to denote systems that used handcrafted or predefined rule sets. As a result, we concentrate on rule-based ML techniques, such as classification and association rule learning techniques.[99] Several popular classification techniques such as decision trees,[48] RIDOR,[100] RIPPER[101] as well as recently proposed IntrudTree,[32] BehavDT[50] and so forth by Sarker et al exist with the ability of rule generation. Based on several metrics, such as support and confidence levels, association rules are created by searching for frequent IF-THEN pattern data. AIS,[60] Apriori,[61] FP-Tree,[62] RARM,[63] Eclat[64] as well as recently proposed ABC-RuleMiner by Sarker et al[65] by taking into account non-redundant generation, are popular association mining techniques. The rule-based modeling paradigm can also be applied to cybersecurity to extract security insights and intelligent decision-making. For example, the IF-THEN rule for identifying anomalies can be expressed as "IF the flag value is *RSTR*, THEN the outcome is *anomaly*". Another rule with many security features may be "IF flag value is *SF*, service is *ftb*, and duration <= 4, THEN the outcome is *anomaly*," which could be created using the tree shown in Figure 7.[32]

Nespoli et al[102] provide a dynamic rule-based system for cyberprotection in IoT contexts. Using rule-based learning, Khorshed et al[103] propose a method for classifying various denial-of-service threats in cloud computing. Holkovivc et al[104] offer a rule-based engine that automates network security analysis at the packet level. Kenaza et al[105] offer a rule-based alert correlation engine in which ontology provides a full environment to represent information for intrusion detection. In addition, a belief rule[106] is an expansion of the conventional IF-THEN rule that uses a belief structure in the consequent part, which is taken into account modeling under uncertainty as well. The belief rule's antecedent portion is made up of one or more antecedent attributes with associated referential values, whereas the consequent portion is made up of one consequent attribute. For instance, a belief rule-based anomaly detection under uncertainty has been presented.[107] Thus we can conclude that rule-based modeling with the variants mentioned above, can play a significant role in the development of AI models as well as intelligent decision-making in various cybersecurity concerns, depending on the characteristics of the problem.

3.5 | Uncertainty and fuzzy logic-based approach

Models built using fuzzy logic may recognize, represent, control, understand, and use data and information that are imprecise and uncertain as fuzzy logic is a precise logic of approximation and imprecision reasoning.[108]

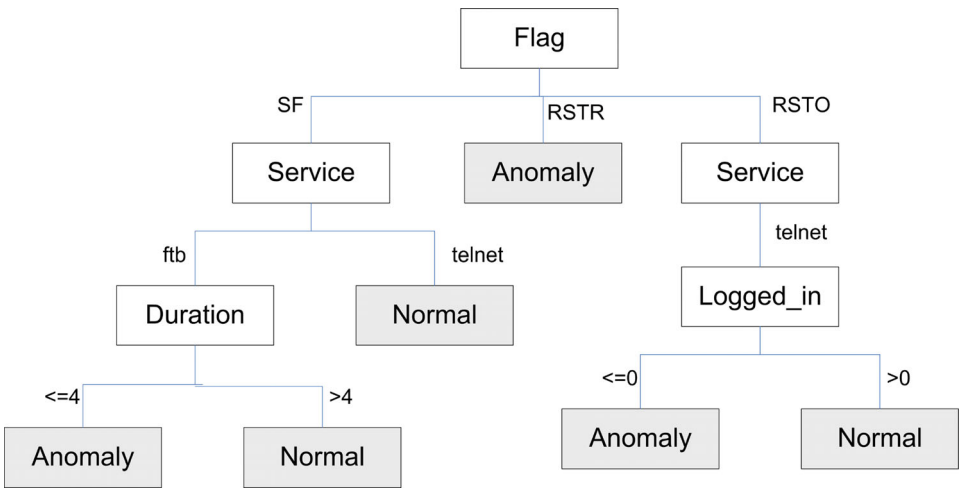

FIGURE 7 An example of detecting cyber-anomalies based on a decision tree-based machine learning model

Fuzzy logic is preferred when differentiating features are vague and depend on human skill and understanding, even if ML models may discriminate between two or more object classes based on their capacity to learn from data. As a result, the system may function with any form of input data, including data that is imprecise, distorted, or noisy, as well as data that is constrained. Because it leverages data collected in environments with such features, it's a good method to use in scenarios with real, continuous-valued elements.[109] Various concerns in cybersecurity are addressed using fuzzy logic-based models. For example, Vorobiev et al[110] perform a fuzzy logic analysis of computer security incidents to examine the detection, prevention, and removal of cyberattack consequences. Alali et al provide a fuzzy logic inference system for upgrading the risk assessment model of cyber security.[111] Joshi et al provide a fuzzy logic-based feature engineering strategy for Botnet identification using ANN in their paper.[112] Saidi et al provide a fuzzy logic-based intrusion detection system as a service for malicious port scanning traffic detection.[113] Haripriya et al[114] describe an effective fuzzy logic-based technique to identify DoS attacks in the MQTT protocol for the internet of things. Overall, we can conclude that fuzzy logic can reach reasonable results in a world of imprecision, vague, distorted, or incomplete data, as well as under uncertain conditions, in some cases belief rule-based modeling mentioned above, could be useful while developing a cyber-model in relevant problem domains.

3.6 | Knowledge representation, uncertainty reasoning, and expert system modeling

Knowledge representation is the study of how to effectively characterize an intelligent agent belief, intentions, and assessments for automated reasoning, where reasoning is the act of making conclusions, predicting outcomes, or coming up with explanations based on knowledge. It has emerged as one of the most promising areas in AI. Logic, semantic network, frame, and production rules are just a few of the knowledge representation methodologies that can be used to build a knowledge-based conceptual model.[115] We discuss potential knowledge representation systems in the following, taking into account real-world concerns.

- *Ontology-based*: Ontologies are conceptual representations of what exists in a domain that has been converted into a machine-readable form using information representation techniques. There are several sorts of ontologies used in the area[115]: top-level ontologies or higher ontologies, domain ontologies, and application ontologies. In general, ontology is "an explicit specification of conceptualization and a formal way to define the semantics of knowledge and data."[116] According to Reference 116, formally, an ontology is represented as "$\{O = C, R, I, H, A\}$, where $\{C = C_1, C_2, \ldots, C_n\}$ represents a set of concepts, and $\{R = R_1, R_2, \ldots, R_m\}$ represents a set of relations defined over the concepts. I represents a set of instances of concepts, H represents a directed acyclic graph defined by the subsumption relation between concepts, and A represents a set of axioms bringing additional constraints on the ontology." A conceptual modeling framework for a cybersecurity system is shown in our earlier paper by Sarker et al[21] that takes security ontologies into account, along with the information flow from data sources to applications, which could be useful in relevant application areas.
- *Rule-base, uncertainty and probabilistic reasoning*: A rule that is frequently stated as an "IF $<$ *condition* $>$ THEN $<$ *action* $>$" statement,[99] also referred to as a conditional statement with a hypothesis and a conclusion or decision, mentioned earlier. The key benefit of a rule-based system is that it allows us to easily add, remove, or update rules according to current needs,[117] which could be an important aspect in the context of cybersecurity. Using the concept of probability to denote the degree of knowing uncertainty could be another method of knowledge representation known as probabilistic reasoning.[118] To deal with uncertainty in a model probabilistic models, fuzzy logic, and belief rules, and similar other strategies could be employed.

Based on the knowledge representation mentioned above, an AI expert system could be built effectively, which is typically a computer program that can make decisions similar to a human expert in a particular problem domain. A cybersecurity expert system is an example of a knowledge-based or rule-based system, which typically consists of two subsystems: an inference engine and a knowledge base defined as security rules.[21] The knowledge base is the core component of this cybersecurity expert framework, as it contains domain knowledge as well as operational knowledge of security decision rules, while the inference engine uses rules to deduce new facts from existing knowledge from a security perspective. In addition to human experts, several techniques, such as classification learning rules, association learning rules, fuzzy logic-based rules, and belief rules as well as conceptual semantic rules, can be used to extract meaningful rules to build a rule-based cybersecurity expert system. Overall, we can infer that, based on its computing capabilities

and knowledge to produce intelligent decisions, cybersecurity expert systems modeling could be an essential aspect of the AI-driven cybersecurity industry.

3.7 | Case-based reasoning

CBR is concerned with the "smart" reuse and adaption of information from previously solved problems ("cases") to new and unsolved challenges. Case-based reasoners handle new problems by retrieving and changing previously stored "cases" that describe similar previous problem-solving experiences. The principle is that the more the two issues are alike, the more similar their solutions will be. In CBR research, the CBR process is explored as a model of human cognition as well as a mechanism for building intelligent systems. CBR is utilized for a variety of applications. Nunes et al[119] provide a CBR technique for cybersecurity event recording and resolution in their paper. Lansley et al[120] provide a CBR approach for detecting social engineering attacks. For digital forensics, Al et al[121] present a CBR approach for evaluating cyber-attack intent. CBR becomes increasingly intelligent as the number of saved examples grows, therefore it could be useful in such scenarios while creating a model. The system's efficiency will, however, decrease as the time it takes to discover and handle relevant cases grows.

3.8 | Text mining and natural language processing

Cybercriminals and security defense tools are increasingly relying on NLP to comprehend and process unstructured data. The ultimate goal of NLP is to extract knowledge from unstructured data or information, that is, to effectively interpret, decipher, comprehend, and make sense of human languages.

When unstructured security information is available, some aspects of NLP, such as lexical analysis, syntactic analysis, and semantic analysis, can be used to model intelligent cybersecurity. For example, lexical analysis of domain names will lead to the construction of an NLP-based model to identify malicious domains that may encompass the "malicious nature" of domains used by cybercriminals.[122] A syntactic analysis, such as parsing[123] could help construct an NLP-based model for cyberattack prediction or extract meaningful data from enormous amounts of public text. Semantic analysis, which includes analyzing the context and perception of words as well as the structure of sentences, is another important way of solving NLP tasks.

Latent semantic analysis with keyword extraction,[124] for example, can be used for phishing classification. Most NLP-based modeling relies on the above-mentioned machine and deep learning techniques[41,69] to create a data-driven analytical model that can be used for a variety of purposes in the cybersecurity domain, including detecting malicious domain names, vulnerability analysis, phishing identification, malware family analysis, and so on. In recent days, NLP-based modeling has played a significant role in a variety of cybersecurity issues, including automatic detection of cyberbullying by analyzing social media text,[125] detecting cybercrime in text-based online interactions,[126] cybersecurity threat awareness from tweets,[127] and many others.

Overall, an NLP-based methodology can be utilized to improve cybersecurity operations by automating threat intelligence derived from unstructured sources. Thus NLP combined with ML approaches is considered a driver for the automation of security tasks due to its security modeling capabilities, which vary depending on the target security application.

3.9 | Visual analytics, computer vision, and pattern recognition

Using digital images, videos, and other visual inputs, computer systems may extract information that can be used to act or make recommendations. Computer vision[128] is a subfield of AI that enables this. It takes an engineering approach to understand and automate the operations that the human visual system is capable of. Thus the ultimate goal of this type of research is to automate the extraction, analysis, and comprehension of relevant information from a single image or a collection of images. It involves technologically manipulating an image down to the pixel level to lay the theoretical and computational groundwork for autonomous visual cognition. In the field of visual analytics and computer vision, popular tasks include object recognition or classification, detection, tracking, and image restoration.

Modern computer vision techniques are built on the principle of pattern recognition, which is the automatic detection of patterns and regularities in data. In pattern recognition, it is common to categorize, that is, supervised learning, and

group, that is, unsupervised learning patterns.[41] Even though ML[41] and DL[69] are more recent approaches to pattern identification, they nevertheless have their roots in statistics and engineering. This is because enormous amounts of data are now more readily available, and there is also an increased amount of processing power.

The concept of such visual analytics can be applied to a variety of cybersecurity issues. For example, Rao et al[129] propose a computing vision technique for detecting phishing attacks that uses a legitimate image database that includes all major website screenshots as well as their URLs. Corum et al[130] propose image visualization and processing technique for detecting PDF malware. Straub et al[131] provide a model for image pattern recognition systems and, more broadly, pattern recognition systems in general, in which they identify various areas of vulnerability and explore the types of attacks, that each of these sites is particularly vulnerable to. Vidal et al[132] describe a pattern recognition technique for detecting Android malware through the analysis of suspicious boot patterns. The authors in Reference 133 created a malware detection system that converts malware files into image representations and uses CNN to classify the image representations. Through their rigorous investigation, Freitas et al[134] provide a large-scale cybersecurity image database of malicious software to the computer vision community, unlocking new and interesting cybersecurity prospects. Kyrkou et al[135] provide AI-based cybersecurity for protecting automated driving systems from camera sensor threats. As a result, such analytics are crucial for developing visual AI models to address a variety of real-world cybersecurity concerns.

3.10 | Hybrid approach, searching, and optimization

A "hybrid approach" is a combination of several approaches or systems used to create a new and better model. As a result, depending on the demands, a hybrid strategy combines the required methodologies listed above. Varzaneh et al[136] for example, provide an intrusion detection system based on a new fuzzy rule-based classification system based on a genetic algorithm. Using CBR and DL, Lansley et al[120] propose a method for detecting social engineering attacks. Botnet detection using ANN and fuzzy logic is presented by Joshi et al.[112] Sarker et al also discussed various ML[41] and DL[69] techniques, as well as their hybridization, that can be used to solve a variety of cybersecurity issues, such as malware analysis, attack behavior, and threat analysis, phishing detection, ransomware detection, and intrusion detection.

To create a useful AI model in this area, incorporating various approaches may be required. In addition, a large number of AI issues can theoretically be resolved by doing a quick search among a wide pool of potential solutions, which would reduce the complexity of the reasoning process. Furthermore, to solve real-world issues, evolutionary computation employs an optimization search technique like genetic algorithms. For instance, a genetic algorithm is used in cybersecurity to detect irregularities in a fog computing environment.[137] A genetic algorithm is also utilized for optimal feature selection to identify Android malware using ML techniques.[138] The platform learns from the data and provides the most accurate and relevant search results automatically with AI-powered search. As a result, while developing AI models to handle real-world problems, searching and optimization strategies can be applied as part of hybridization.

Overall, we can conclude that the 10 categories of potential AI techniques outlined above could be very beneficial in the development of various AI models as well as intelligent decision-making, depending on the nature of the problem and the intended cyber application. In the next section, we will discuss defending robustly against various adversarial attacks, which is still an open question in the context of cybersecurity.

4 | ADVERSARIAL MACHINE LEARNING IN CYBERSECURITY

Adversarial ML is typically defined as "the study of how machine learning algorithms and models, discussed in earlier Section 3, are attacked and how they are defended." For example, network intrusion could be one type of adversarial attack where attackers actively scan the systems to find new network device vulnerabilities and exploit susceptible hosts or systems. Figure 8 illustrates how adversarial attacks can be broadly divided into two types: *poisoning attacks*, in which the attacker modifies the training data or its labels to make the model perform poorly during deployment, and *evasion attacks*, in which the attacker manipulates the data during deployment to deceive previously trained classifiers. Thus adversarial ML in the context of cybersecurity may have two primary branches, one of which actively develops new ways to attack already-in-use ML algorithms and systems, and the other of which attempts to greatly enhance the resilience of ML approaches against attacks. To effectively handle the adversarial samples produced with only minor perturbations, ML techniques should then significantly enhance their generalization capability. The development of secure, resilient ML systems that can withstand numerous forms of malicious attacks is a crucial and urgent endeavor as well as a significant research issue of today's context as well.

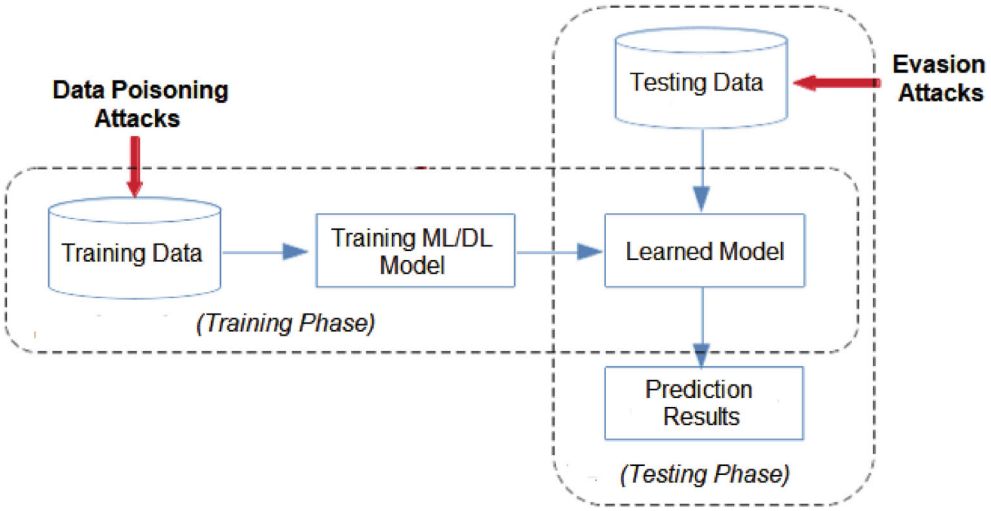

FIGURE 8 Illustration of adversarial attacks on train and test data with machine learning modeling

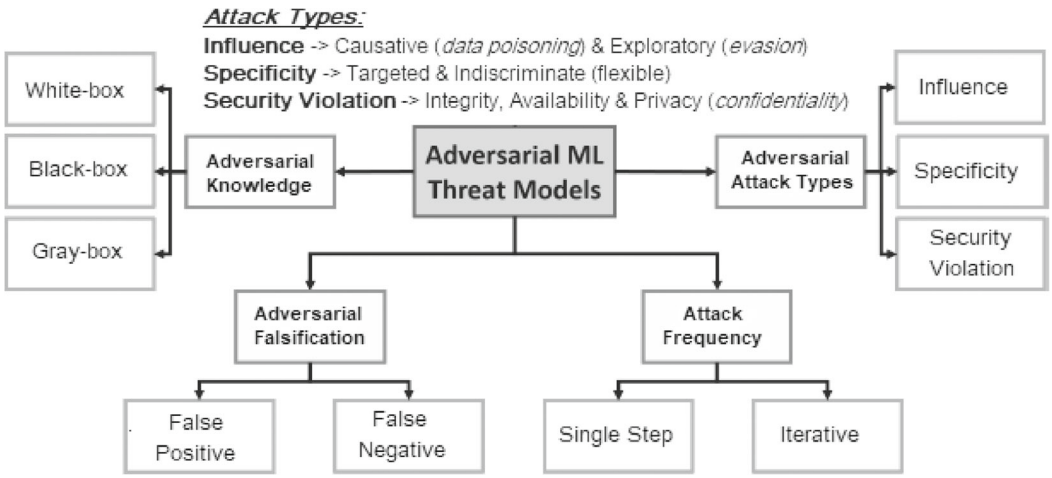

FIGURE 9 The taxonomy of various types of threat models used in literature to design adversarial ML attacks

4.1 | Adversarial threat model

A key element of security analysis is threat modeling. The main objective of threat modeling is to maximize the system's security through the establishment of security objectives, the identification of potential threats and vulnerabilities, and the creation of countermeasures for preventing or minimizing the consequences on the system of those threats. In the following, we discuss the relevant aspects of the adversarial ML threat shown in Figures 8 and 9.

Huang et al[37] proposed a formal taxonomy to model an adversary by taking into account three aspects as influence, security violation, and specificity. Influence comprises causative attacks which aim to gain influence over training data, and exploratory attacks that use different methods, including probing the detector, to learn more about the target or its training data rather than altering the training process. Integrity, availability, confidentiality, or privacy violation is all indicators of a security breach, shown in Figure 1 as well. Attacks on availability lead to so many classification errors, including false positives and false negatives that the system is practically rendered worthless. Intrusion points are labeled as normal as a result of integrity attacks that is defined as *false negatives*, which could be considered a serious research issue in the context of cybersecurity. A privacy violation occurs when an adversary gets information from a learner, compromising the users' confidentiality or privacy. Specificity involves indiscriminate and targeted attacks on a particular occurrence. In a targeted attack, only one or a limited number of target points are the main emphasis. A more flexible object, such as any false negative, is the focus of an indiscriminate adversary.

In addition to such attack types mentioned above, adversarial attacks are further classified into three categories such as white-box, gray box, and black box based on the adversarial knowledge. While white-box attacks presume complete knowledge of the underlying ML model including the optimization technique, model architecture, hyperparameters and so forth black-box adversarial attacks presume the adversary has no knowledge and no access to the underlying ML model or the training data. Gray-box attacks assume that the adversary has some knowledge of the targeted model. False positive attacks and false negative attacks are the two different types of falsification attacks that the adversary can launch. An adversary attempts false positive attacks by producing a negative sample that could be mistaken for a positive one. On the other hand, when making an effort at a false negative attack, the adversary creates a positive sample that can be mistaken for a negative one. This kind of attack is referred to as an evasive attack in adversarial ML. A single step or an iterative optimization procedure may be used in adversarial attacks. Iterative adversarial attacks are more potent than single-step attacks, but they demand a lot of time and computational resources to generate efficiently since they include several interactions with the ML system. In Figure 9, we have shown taxonomy of adversarial threat models by taking into account all the above-discussed factors.

Depending on how the attack algorithm creates the perturbation, adversarial attacks can also be structured. An adversarial attack may employ one of the following strategies: (i) gradient; (ii) transferability or score; (iii) decision; or (iv) approximation.[42] The gradient-based algorithms make use of comprehensive information about the gradient of the target model relative to the input that is provided. This method of attack is typically used in white-box settings where the attacker has complete access to and knowledge of the targeted model; the transfer or score-based attack algorithm is frequently helpful in black-box attacks that either rely on gaining access to the dataset utilized by the intended model or the scores anticipated by it to roughly estimate a gradient. Since decision-based attacks involve fewer parameter changes than gradient-based attacks, they are thought to be easier and more flexible. To feed conventional gradient-based attacks, approximation-based attacks employ algorithms that use a differentiable function to approximatively construct the outputs of a non-differentiable or randomized layer of either a model or defense.

4.2 | Adversarial defense methods

Defending robustly against adversarial attacks is still an open question. Defenses should always include a statement in their threat model that they are strong against adversarial attacks. Defense methods are designed based on two main approaches.[39,139]

- *Proactive defenses*: Before the attack, the target system is set up to handle foreseeable threats. By incorporating knowledge of the attacker's intention into the defense mechanism or developed algorithm, they strive to prevent potential attacks or mitigate their effects before they materialize. In a proactive approach, the designer should also make an effort to predict the adversary's behaviors by determining the most pertinent threats, coming up with effective countermeasures if needed, and repeating this process before system deployment.
- *Reactive defenses*: Defense strategies are used in response to an attack, such as machine unlearning following poisoning attacks. They are only updated when new attacks are discovered, and then only by adding new features if necessary or, more frequently, by retraining the system using the most recent data. Reactive strategies do not attempt to predict upcoming attacks or forecast future security weaknesses, leaving the system exposed to them. In addition to a pure proactive approach focused on reducing the risk of future attacks mentioned above, system security can also be improved reactively by reflecting on the past and learning from it, which is sometimes more efficient and practical.

In general, the proactive security method aids in predicting cyberattacks in advance, whereas the reactive security approach focuses on minimizing the damage done by cybersecurity threats and addressing vulnerabilities. The majority of defenses employ proactive defense strategies to limit the damage as much as possible.[140] The proactive defense methods can be implemented by *modifying the ML model*, where adversarial sample thwarting techniques like data transformation, noise filtering, or mapping to normal samples, training process alteration techniques like adversarial training or altering the training process, and ML algorithm modification techniques like applying non-linear ML algorithms and designing robust ML algorithms are commonly used. Another proactive defensive strategy might be to apply a *specialized detector*, such as feature squeezing or applying successive detectors.[39] For greater defense gain, different defense strategies can also be integrated (also known as ensemble defenses), but this method is not always more effective.

According to Rosenberg et al,[38] "adversarial defense methods can be categorized as detection and robustness methods, where *detection methods* are used to detect adversarial examples and *robustness methods* are used to enhance a classifier's rigidity to them without explicitly attempting to detect them." Every defense strategy is either attack-specific, requiring adversarial examples produced by the attack to mitigate the attack, or attack-agnostic, effective against all attack strategies without requiring a dataset of adversarial examples produced by those attacks. Examples of works in the field of cybersecurity where an attack-agnostic method has been employed include the development of an Android malware classifier that is robust to poisoning attacks,[141] evaluation of multiple attack-agnostic robustness defense methods including an ensemble of classifiers,[142] and detection of adversarial attacks for the defense of cyber-physical systems.[143] Similar to this, an attack-specific robustness defense strategy has been used in adversarial retraining against RNN classifiers for spam detection[144] and anomaly detection system of sensor data.[145] As attack-agnostic defense strategies are more general, these should be preferred and the focus of further research to solve real-world issues in the context of cybersecurity.

Overall, this area of cyber security is particularly intriguing because it is rife with adversaries, such as malware creators trying to bypass next-generation antivirus software, intrusion prevention, and other systems based on machine and DL. Adversarial learning in the context of cyber security is nearly identical to the cat-and-mouse game played in the domain of traditional cyber security, where the attackers utilize more sophisticated techniques to escape the defenders and vice versa.[38] Most current research on adversarial ML is supervised learning-focused.[146] The labeling of numerous data points or samples from the most recent attacks, however, may necessitate the use of expensive human knowledge and end up becoming a substantial bottleneck. How to identify adversarial data in unsupervised and weakly supervised settings requires greater focus. It's crucial to quantify the robustness and accuracy trade-off for adversarial ML algorithms in the interim. Even though certain robustness or uncertainty measurements have been put forth in recent works, more investigation into the trade-off is necessary to create resilient learning algorithms.

5 | REAL-WORLD APPLICATIONS WITH RESEARCH ISSUES

In this section, we have highlighted several potential application areas in the context of cybersecurity with research issues. Our main focus is on how different AI technologies discussed earlier might enable security intelligence as well as the development of automated, intelligent, and robust cyber solutions in various application areas as depicted in Figure 10. We have also listed several AI tasks and procedures in Table 1 that are used to solve a variety of real-world cybersecurity issues. Thus, cybersecurity intelligence and robustness modeling as well as automated decision-making using AI techniques in real-world cyber domains have a wide range of future possibilities. While our study has established a solid foundation for cyber intelligence and robustness through multi-aspects AI-based modeling discussed briefly throughout the article, we have identified several research concerns below to conduct further research and development.

- Real-world cyber problems can be resolved using a variety of potential AI methods, such as learning methods, analytics, knowledge discovery methods and so forth depending on the characteristics of data, as discussed broadly in Section 3. To identify an appropriate solution, for example, detecting cyber-anomalies or multi-attacks,[20] it is necessary to comprehend the nature of the cyber problem and data as well as to conduct an in-depth investigation. Thus the question is—"Which AI technique or their ensembles is most suitable to address a certain real-world cybersecurity issue, e.g., cyber incident response, given the circumstances of the situation?"
- The most crucial task for solving real-world cybersecurity issues is a general framework supporting AI-based modeling. Thus the development of such an effective framework that can handle a certain real-world cyber issue could be considered one of the important research directions nowadays. A well-designed security framework ensuring effectiveness and efficiency through rigorous experimental analysis is essential to tackle the target cyber issues to overcome. Therefore the question is—"Which security aspects need to be taken into account while designing an optimal AI-based framework to achieve the desired outcome?"
- Nowadays, our technology-dependent world is a rich source of cybersecurity data, such as anomaly data, fraud data, malware data and so forth[11] due to the increasing popularity of the Internet of Things and digitization. Understanding the nature of cybersecurity data, which includes various types of data breaches and pertinent information, is vital. The idea is to leverage raw security data acquired from pertinent cyber sources to build a data-driven security model that will help us accomplish our target by analyzing the numerous patterns of security incidents or malicious behavior. Various ML and knowledge discovery techniques can be employed to extract such patterns or knowledge. Thus the

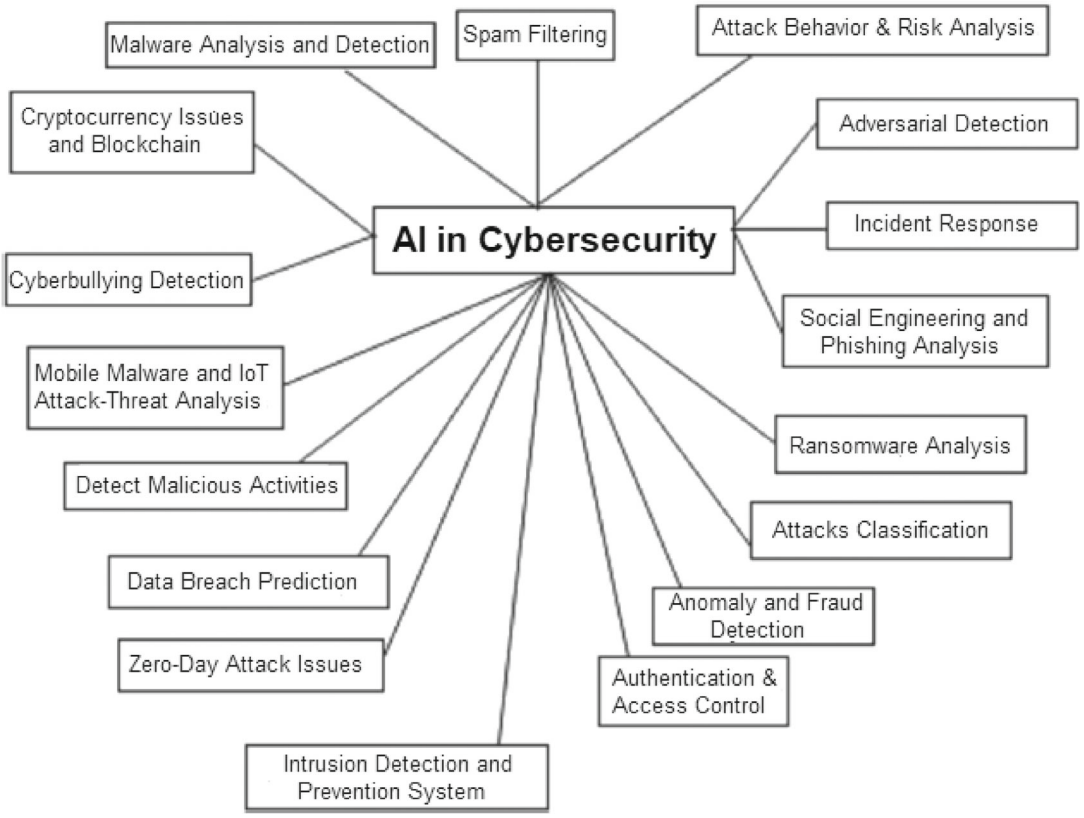

FIGURE 10 Several potential real-world application areas of artificial intelligence (AI) in the context of cybersecurity

question is—"How effectively we can extract insights or useful information from cyber data and use that knowledge to create an automated, intelligent model in the context of cybersecurity??"

- In many cases within the context of cybersecurity, the conventional ML[41] and DL[69] techniques as well as other AI methods discussed broadly in Section 3, may not necessarily be precisely applicable to get the expected outcome. Designing new methods or their variants by taking into account model optimization, accuracy, and applicability depending on the intended cyber problem as well as relevant data characteristics could be a novel contribution. Thus the question is, "How can we design an effective learning algorithm or AI method for building a security model to get the expected outcome for a particular cyber issue?"
- As mentioned earlier, the ML[41] and DL[69] methods as well as knowledge discovery and several other AI methods, typically learn from data. The outcomes of any analytical algorithm are directly impacted by data preprocessing techniques such as data cleaning, and transforming, as well as ensuring data quality for modeling. To ensure that the data behave properly, it is crucial to carry out particular data preprocessing operations before feeding the data into the eventual AI model.[12] Therefore the question is—"How to effectively represent and feed cyber data to a machine learning or AI model by ensuring data quality for a better outcome in a particular cyber issue?"
- In the field of AI, decision rules, expressed as an IF-THEN statement, could be very useful to solve various real-world problems including cybersecurity issues.[99] A cyber expert system, whose knowledge base includes security policy rules, is an example of a rule-based system that enables decisions to be made following security policies. In addition, a belief rule[106] is an expansion of the conventional IF-THEN rule that uses a belief structure in the consequent part while generating policy rules. A rule-based system may therefore effectively manipulate knowledge and conduct data analysis. Thus, the question is—"How can we build an automated rule-based system that mimics the decision-making capabilities of a human expert by effectively discovering a concise set of IF-THEN rules from relevant cyber data?"
- Uncertainty is an integral part of the risk, which typically refers to the lack of confidence or certainty in an occurrence, such as information obtained through untrustworthy sources. Several techniques such as the probability-based model, fuzzy logic as well as belief rules, addressed in Section 3 allow for the processing of uncertain and imprecise knowledge while simultaneously providing a sophisticated reasoning framework. AI-based security models should be able

TABLE 1 A summary of AI tasks and methods to solve various cybersecurity issues

AI techniques	Cybersecurity tasks	References
Machine learning	Intrusion detection analysis	Lin et al[147]
	Cyber learning with attack and anomaly detection	Sarker et al[20]
	Detecting malicious activities and intrusions	Alauthman et al[68]
	Malicious behavior analysis	Moon et al[148]
	Denial of service attack detection	Syed et al[149]
Neural network and deep learning	IoT botnet traffic classification	Javed et al[150]
	Malicious user detection	Hong et al[151]
	Detecting phishing websites	Xiao et al[152]
	Time-based botnet detection	Shi et al[153]
	Zero-day malware detection	Kim et al[91]
Data mining, knowledge discovery, and advanced analytics	Cybersecurity data science framework	Sarker et al[11]
	Detecting cyber-attacks	Tianfield et al[97]
	Association analysis of cyber-attack attribution	Li et al[98]
Rule-based modeling and decision-making	To defend smart devices in IoT environments	Nespoli et al[102]
	Classifying different denial-of-service attacks in cloud computing	Khorshed et al[103]
	An automating network security analysis at packet-level	Holkovivc et al[104]
	A rule-based engine for security alert correlation	Kenaza et al[105]
	A belief rule-based anomaly detection	Ul et al[107]
Fuzzy logic-based approach	Analysis of computer security incidents	Vorobiev et al[110]
	Improving risk assessment model of cyber security	Alali et al[111]
	Botnet detection	Joshi et al[112]
	Intrusion detection system	Saidi et al[113]
	To detect DoS attack in MQTT protocol for Internet of Things	Haripriya et al[114]
Knowledge representation, uncertainty reasoning, and expert system modeling	Designing an ontology-based cybersecurity framework	Mozzaquatro et al[154]
	An ontology-based security risk management model	Arogundade et al[155]
	Knowledge-based cybersecurity expert system modeling	Sarker et al[21]
Case-based reasoning	Cybersecurity incident recording and resolution	Nunes et al[119]
	Detecting social engineering attacks	Lansley et al[120]
	Analyzing cyber-attack intention for digital forensics	Al et al[121]
Text mining and natural language processing	To classify the malicious domains	Kidmose et al[122]
	Cybersecurity threat awareness from tweets	Alves et al[127]
	Detecting cybercrime in text-based online interactions	Sekeres et al[126]
	Automatic detection of cyberbullying through analyzing social media text	Van et al[125]
Visual analytics, computer vision and pattern recognition	A computer vision technique to detect phishing attacks	Rao et al[129]
	A robust PDF malware detection with image visualization techniques	Corum et al[130]
	A model for image pattern recognition system considering security issues	Straub et al[131]
	In automated malware detection from large-scale image database	Freitas et al[134]
	Combat security risks against camera sensor attacks	Kyrkou et al[135]
Hybrid approach, searching, and optimization	Intrusion detection using a fuzzy rule and genetic algorithm	Varzaneh et al[136]
	Detecting social engineering attacks using CBR and deep learning	Lansley et al[120]
	Botnet detection using ANN and fuzzy logic	Joshi et al[112]

to comprehend and mitigate uncertainty and risk to be used to solve decision-making problems. Thus the question is—"How to effectively handle uncertainty in an AI-enabled decision-making system in the context of cybersecurity."

- When it comes to cybersecurity, ML techniques[41] are frequently used to identify problems where adversaries deliberately change their objects to evade detection. One sort of adversarial assault is a network incursion, in which attackers actively search the systems for new network device vulnerabilities and target vulnerable hosts or systems. Thus, the study of adversarial ML is important, discussed briefly in Section 4, which has two main branches, one of which actively creates new ways to attack ML algorithms and systems that are currently in use, and the other of which aims to significantly increase the robustness of ML approaches against attacks. To effectively handle the adversarial samples, ML techniques should then significantly enhance their generalization capability. Thus the development of secure, resilient ML systems that can withstand numerous forms of malicious attacks is a crucial and urgent endeavor as well as a significant research issue in the area. Therefore the question is—"How to design robust algorithms or models to effectively handle such adversarial issues in today's cybersecurity?"
- Context is typically described as any information that can be used to characterize the situation of an entity and used for various purposes.[156-158] Based on this, a security context could be any information such as temporal, spatial, relationship, connectivity, activity, individuality and so forth that can be used to characterize the security situation of a cybersecurity environment. Such contextual information can improve the effectiveness of security decisions, for example, intrusion detection precision, as well as the efficiency and flexibility of the resource management process. Thus it's important to understand contextual information categories and modeling to effectively use security contexts in a cybersecurity environment. Thus the question is—"How to effectively incorporate security contexts in an AI-enabled context-aware model to adapt their behavior dynamically in a real-world cyber space?"
- Cyber resilience is the capability to defend electronic information and systems against cyber-attacks and to quickly resume commercial activities in the event of a successful attack. Businesses that have strong cyber resilience understand that attackers may have access to cutting-edge tools, zero-day vulnerabilities, and the element of surprise. This concept assists organizations in preparing for, preventing, responding to, and successfully recovering from their pre-attack business operations and processes. To be more resilient to attacks, an organization or company needs to change the way it thinks and become more agile. Thus the question is—"How to generalize and design such a security framework handling relevant cyber resilience issues?"

Overall, cybersecurity intelligence and robustness could be considered as a wide-open area where academics can contribute by proposing new techniques or enhancing those that already exist to address the aforementioned issues and deal with diverse cyber issues in a range of application areas. This can also assist the researchers in doing a complete investigation of the application's hidden and unforeseen difficulties to produce results that are more trustworthy and realistic. Overall, we can conclude that addressing the aforementioned problems and contributing to the development of effective and efficient techniques may result in the modeling of "Future Generation Cybersecurity" which is based on "intelligence" and "robustness" as well as automated decision-making focusing in this article in the cybersecurity domain.

6 | CONCLUSION

The growing significance of cybersecurity and AI motivated us to present in this article a comprehensive view of cybersecurity intelligence and robustness, taking into account AI-based modeling to address various security challenges. Our main focus was on how different AI technologies might enable security intelligence as well as the development of automated, intelligent, and robust solutions, depending on the nature and diversities of cybersecurity issues. It begins with research motivation, moves on to widespread AI approaches including adversarial ML, and then makes advancements in several of today's diverse cybersecurity concerns. Each key strategy has been explored in terms of significant security research as well as how it might be applied to intelligently tackle current cybersecurity issues. Such AI-based modeling can be used in a range of problem domains, from detecting attacks to predicting risky behavior that could result in a phishing attack or malicious code, all of which are briefly covered in this article. We have also explored the adversarial threat model as well as different defense methods against adversarial attacks. We then highlighted several significant security analysis concerns to provide a direction for future research in the field of cybersecurity intelligence. Since AI offers a lot of potential for applications in cyber security, the researcher and professional communities need to be aware of the current state-of-the-art and the related issues, discussed in this article. Overall, we believe that our exploration into

cybersecurity intelligence and robustness through multi-aspects AI-based modeling is on the proper track and might serve as a reference guide for academia and industry professionals conducting future-generation cybersecurity research and real-world applications.

ACKNOWLEDGMENT

Open access publishing facilitated by Edith Cowan University, as part of the Wiley - Edith Cowan University agreement via the Council of Australian University Librarians. [Correction added on 13 January 2023, after first online publication: CAUL funding statement has been added.]

CONFLICT OF INTEREST

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Iqbal H. Sarker https://orcid.org/0000-0003-1740-5517

REFERENCES

1. Sun N, Zhang J, Rimba P, Gao S, Zhang LY, Xiang Y. Data-driven cybersecurity incident prediction: a survey. *IEEE Commun Surv Tutor*. 2018;21(2):1744-1772.
2. McIntosh T, Jang-Jaccard J, Watters P, Susnjak T. The inadequacy of entropy-based ransomware detection. Proceedings of the International Conference on Neural Information Processing; 2019:181–189; Springer.
3. Alazab M, Venkatraman S, Watters P, Alazab M. Zero-day malware detection based on supervised learning algorithms of API call signatures. Proceedings of the Ninth Australasian Data Mining Conference; 2010.
4. Shaw A. Data breach: from notification to prevention using PCI DSS. *Columbia J Law Soc Probl*. 2009;43:517-562.
5. Gupta BB, Tewari A, Jain AK, Agrawal DP. Fighting against phishing attacks: state of the art and future challenges. *Neural Comput Appl*. 2017;28(12):3629-3654.
6. Balakrishnan V, Khan S, Arabnia HR. Improving cyberbullying detection using twitter users' psychological features and machine learning. *Comput Secur*. 2020;90:101710.
7. Sarker IH, Khan AI, Abushark YB, Alsolami F. Internet of Things (IoT) security intelligence: a comprehensive overview, machine learning solutions and research directions. *Mobile Netw Appl*. 2022;1-17.
8. IBM Security Report. https://www.ibm.com/security/data-breach. Accessed October 20, 2019.
9. Fischer EA. Cybersecurity issues and challenges: in brief; 2014.
10. Juniper Research. https://www.juniperresearch.com/. Accessed October 20, 2019.
11. Sarker IH, Kayes A, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. *J Big Data*. 2020;7(1):1-29.
12. Sarker IH. Smart city data science: towards data-driven smart cities with open research issues. *Internet Things*. 2022;19:100528.
13. Anwar S, Mohamad Zain J, Zolkipli MF, et al. From intrusion detection to an intrusion response system: fundamentals, requirements, and future directions. *Algorithms*. 2017;10(2):39.
14. Mohammadi S, Mirvaziri H, Ghazizadeh-Ahsaee M, Karimipour H. Cyber intrusion detection by combined feature selection algorithm. *J Inf Secur Appl*. 2019;44:80-88.
15. Tapiador JE, Orfila A, Ribagorda A, Ramos B. Key-recovery attacks on KIDS, a keyed anomaly detection system. *IEEE Trans Dependable Secur Comput*. 2013;12(3):312-325.
16. Tavallaee M, Stakhanova N, Ghorbani AA. Toward credible evaluation of anomaly-based intrusion-detection methods. *IEEE Trans Syst, Man, Cybern C*. 2010;40(5):516-524.
17. Foroughi F, Luksch P. Data science methodology for cybersecurity projects. arXiv preprint arXiv:1803.04219, 2018.
18. Saxe J, Sanders H. *Malware Data Science: Attack Detection and Attribution*. San Francisco, CA: No Starch Press; 2018.
19. Sarker IH. AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN Comput Sci*. 2022;3:158.
20. Sarker IH. CyberLearning: effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. *Internet Things*. 2021;14:100393.
21. Sarker IH, Furhad MH, Nowrozy R. AI-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Comput Sci*. 2021;2:173.
22. Sarker IH. Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. *Ann Data Sci*. 2022.
23. Rainie L, Anderson J, Connolly J. Cyber attacks likely to increase. Digital Life in vol. 2025, 2014.

24. Al-Garadi MA, Mohamed A, Al-Ali AK, Du X, Ali I, Guizani M. A survey of machine and deep learning methods for Internet of Things (IoT) security. *IEEE Commun Surv Tutor*. 2020;22(3):1646-1685.
25. Craigen D, Diakun-Thibault N, Purse R. Defining cybersecurity. *Technol Innov Manag Rev*. 2014;4(10):13-21.
26. Aftergood S. Cybersecurity: the cold war online. *Nature*. 2017;547(7661):30-31.
27. National Research Council. Toward a safer and more secure cyberspace; 2007.
28. Jang-Jaccard J, Nepal S. A survey of emerging threats in cybersecurity. *J Comput Syst Sci*. 2014;80(5):973-993.
29. Maalem Lahcen RA, Caulkins B, Mohapatra R, Kumar M. Review and insight on the behavioral aspects of cybersecurity. *Cybersecurity*. 2020;3:1-18.
30. Dai J, Chen C, Li Y. A backdoor attack against LSTM-based text classification systems. *IEEE Access*. 2019;7:138872-138878.
31. Wang B, Yao Y, Shan S, et al. Neural cleanse: identifying and mitigating backdoor attacks in neural networks. Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP); 2019:707–723; IEEE.
32. Sarker IH, Abushark YB, Alsolami F, Khan AI. IntruDTree: a machine learning based cyber security intrusion detection model. *Symmetry*. 2020;12(5):754.
33. Alsayed A, Bilgrami A. E-banking security: internet hacking, phishing attacks, analysis and prevention of fraudulent activities. *Int J Emerg Technol Adv Eng*. 2017;7(1):109-115.
34. Blaise A, Bouet M, Conan V, Secci S. Detection of zero-day attacks: an unsupervised port-based approach. *Comput Netw*. 2020;180:107391.
35. Ekramifard A, Amintoosi H, Seno AH, Dehghantanha A, Parizi RM. A systematic literature review of integration of blockchain and artificial intelligence. In: Choo KK, Dehghantanha A, Parizi R, eds. *Blockchain Cybersecurity, Trust and Privacy*. Cham: Springer; 2020:147-160.
36. Wu J, Liu J, Zhao Y, Zheng Z. Analysis of cryptocurrency transactions from a network perspective: an overview. *J Netw Comput Appl*. 2021;190:103139.
37. Huang L, Joseph AD, Nelson B, Rubinstein BI, Tygar JD. Adversarial machine learning. Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence; 2011:43–58.
38. Rosenberg I, Shabtai A, Elovici Y, Rokach L. Adversarial machine learning attacks and defense methods in the cyber security domain. *ACM Comput Surv*. 2021;54(5):1-36.
39. Sadeghi K, Banerjee A, Gupta SK. A system-driven taxonomy of attacks and defenses in adversarial machine learning. *IEEE Trans Emerg Top Comput Intell*. 2020;4(4):450-467.
40. Sarker IH. Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Comput Sci*. 2021;2:377.
41. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci*. 2021;2(3):1-21.
42. Machado GR, Silva E, Goldschmidt RR. Adversarial machine learning in image classification: a survey toward the defender's perspective. *ACM Comput Surv*. 2021;55(1):1-38.
43. Dua S, Du X. *Data Mining and Machine Learning in Cybersecurity*. Boca Raton, FL: Auerbach Publications; 2016.
44. John GH, Langley P. *Estimating Continuous Distributions in Bayesian Classifiers*. Burlington, MA: Morgan Kaufmann Publishers Inc; 1995:338-345.
45. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. *Mach Learn*. 1991;6(1):37-66.
46. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Comput*. 2001;13(3):637-649.
47. Quinlan JR. Induction of decision trees. *Mach Learn*. 1986;1(1):81-106.
48. Quinlan JR. *C4.5: Programs for Machine Learning*. Burlington, MA: Morgan Kaufmann; 1993.
49. Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*. New York: CRC Press; 1984.
50. Sarker IH, Colman A, Han J, Khan AI, Abushark YB, Salah K. BehavDT: a behavioral decision tree learning to build user-centric context-aware predictive model. *Mobile Netw Appl*. 2019;25:1151-1161.
51. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.
52. Han J, Pei J, Kamber M. *Data Mining: Concepts and Techniques*. Amsterdam, Netherlands: Elsevier; 2011.
53. MacQueen J. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA; vol. 1, 1967:281–297.
54. Rokach L. A survey of clustering algorithms. In: Maimon O, Rokach L, eds. *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer; 2010:269-298.
55. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; vol. 96, 1996:226–231.
56. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
57. Sneath PH. The application of computers to taxonomy. *J Gen Microbiol*. 1957;17(1):201-226.
58. Sorensen T. Method of establishing groups of equal amplitude in plant sociology based on similarity of species. *Biol Skr*. 1948;5.
59. Sarker IH, Colman A, Kabir MA, Han J. Individualized time-series segmentation for mining mobile phone user behavior. *Comput J*. 2018;61(3):349-368.
60. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. *ACM Sigmod Rec*. 1993;22:207-216.
61. Agrawal R, Srikant R. Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases; vol. 1215, 1994:487–499.
62. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. *ACM Sigmod Rec*. 2000;29:1-12.

63. Das A, Ng WK, Woon YK. Rapid association rule mining. Proceedings of the 10th International Conference on Information and Knowledge Management; 2001:474–481; ACM.
64. Zaki MJ. Scalable algorithms for association mining. *IEEE Trans Knowl Data Eng*. 2000;12(3):372-390.
65. Sarker IH, Kayes A. ABC-RuleMiner: user behavioral rule-based machine learning method for context-aware intelligent services. *J Netw Comput Appl*. 2020;168:102762.
66. Bellman R. A Markovian decision process. *J Math Mech*. 1957;6:679-684.
67. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. *J Artif Intell Res*. 1996;4:237-285.
68. Alauthman M, Aslam N, Al-kasassbeh M, Khan S, Al-Qerem A, Choo KKR. An efficient reinforcement learning-based botnet detection approach. *J Netw Comput Appl*. 2020;150:102479.
69. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci*. 2021;2(6):1-20.
70. Xin Y, Kong L, Liu Z, et al. Machine learning and deep learning methods for cybersecurity. *IEEE Access*. 2018;6:35365-35381.
71. Ferrag MA, Maglaras L, Moschoyiannis S, Janicke H. Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. *J Inf Secur Appl*. 2020;50:102419.
72. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86(11):2278-2324.
73. Dupond S. A thorough review on the current advance of neural network structures. *Annu Rev Control*. 2019;14:200-230.
74. Mandic D, Chambers J. *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. New York: Wiley; 2001.
75. Susilo B, Sari RF. Intrusion detection in IoT networks using deep learning algorithm. *Information*. 2020;11(5):279.
76. Kim J, Kim J, Thu HLT, Kim H. Long short term memory recurrent neural network classifier for intrusion detection. Proceedings of the 2016 International Conference on Platform Technology and Service (PlatCon); 2016:1–5; IEEE.
77. Yan J, Qi Y, Rao Q. Detecting malware with an ensemble method based on deep neural network. *Secur Commun Netw*. 2018;2018:7247095.
78. McLaughlin N, Rincon M, Kang B, et al. Deep android malware detection. Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy; 2017:301–308.
79. Vinayakumar R, Soman K, Poornachandran P. Deep android malware detection and classification. Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 2017:1677–1683; IEEE.
80. Adebowale MA, Lwin KT, Hossain MA. Intelligent phishing detection scheme using deep learning algorithms. *J Enterp Inf Manag*. 2020.
81. Tran D, Mac H, Tong V, Tran HA, Nguyen LG. A LSTM based framework for handling multiclass imbalance in DGA botnet detection. *Neurocomputing*. 2018;275:2401-2413.
82. Da'u A, Salim N. Recommendation system based on deep learning methods: a systematic review and new directions. *Artif Intell Rev*. 2020;53(4):2709-2748.
83. Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Trans Signal Inf Process*. 2014;3:E2.
84. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems; 2014:2672–2680.
85. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning*. Vol 1. Cambridge: MIT press; 2016.
86. Marlin B, Swersky K, Chen B, Freitas N. Inductive principles for restricted Boltzmann machine learning. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics; 2010:509–516; JMLR Workshop and Conference Proceedings.
87. Kohonen T. The self-organizing map. *Proc IEEE*. 1990;78(9):1464-1480.
88. Hinton GE. Deep belief networks. *Scholarpedia*. 2009;4(5):5947.
89. Liu L, De Vel O, Chen C, Zhang J, Xiang Y. Anomaly-based insider threat detection using deep autoencoders. Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW); 2018:39–48; IEEE.
90. Qu F, Zhang J, Shao Z, Qi S. An intrusion detection model based on deep belief network. Proceedings of the 2017 VI International Conference on Network, Communication and Computing; 2017:97–101.
91. Kim JY, Bu SJ, Cho SB. Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders. *Inform Sci*. 2018;460:83-102.
92. Khan FA, Gumaei A, Derhab A, Hussain A. A novel two-stage deep learning model for efficient network intrusion detection. *IEEE Access*. 2019;7:30373-30385.
93. Parra GDLT, Rad P, Choo KKR, Beebe N. Detecting Internet of Things attacks using distributed deep learning. *J Netw Comput Appl*. 2020;163:102662.
94. Pericherla S, Ilavarasan E. Transformer network-based word embeddings approach for autonomous cyberbullying detection. *Int J Intell Unmanned Syst*. 2021.
95. Wu Z, Zhang H, Wang P, Sun Z. RTIDS: a robust transformer-based approach for intrusion detection system. *IEEE Access*. 2022;10:64375-64387.
96. Zhao Z, Niu W, Zhang X, Zhang R, Yu Z, Huang C. Trine: syslog anomaly detection with three transformer encoders in one generative adversarial network. *Appl Intell*. 2022;52(8):8810-8819.
97. Tianfield H. Data mining based cyber-attack detection. *Syst Simul Technol*. 2017;13(2):90-104.
98. Li Q, Yang Z, Jiang Z, Liu B, Fu Y. Association analysis of cyber-attack attribution based on threat intelligence. Proceedings of the 2017 2nd Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC 2017); vol. 49, 2017. doi:10.2991/limec‐17.2017
99. Sarker IH, Colman A, Han J, Watters P. *Context-Aware Machine Learning and Mobile Data Analytics: Automated Rule-Based Services with Intelligent Decision-Making*. Cham: Springer; 2022.

100. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA: Morgan Kaufmann; 2005.
101. Witten IH, Frank E, Trigg LE, Hall MA, Holmes G, Cunningham SJ. Weka: practical machine learning tools and techniques with Java implementations; 1999.
102. Nespoli P, Díaz-López D, Mármol FG. Cyberprotection in IoT environments: a dynamic rule-based solution to defend smart devices. *J Inf Secur Appl*. 2021;60:102878.
103. Khorshed MT, Ali AS, Wasimi SA. Classifying different denial-of-service attacks in cloud computing using rule-based learning. *Secur Commun Netw*. 2012;5(11):1235-1247.
104. Holkovič M, Ryšavỳ O, Dudek J. Automating network security analysis at packet-level by using rule-based engine. Proceedings of the 6th Conference on the Engineering of Computer Based Systems; 2019:1–8.
105. Kenaza T. An ontology-based modelling and reasoning for alerts correlation. *Int J Data Min, Model Manag*. 2021;13(1–2):65-80.
106. Zhou ZJ, Hu GY, Hu CH, Wen CL, Chang LL. A survey of belief rule-base expert system. *IEEE Trans Syst, Man, Cybern Syst*. 2019;51(8):4944-4958.
107. Ul Islam R, Hossain MS, Andersson K. A novel anomaly detection algorithm for sensor data under uncertainty. *Soft Comput*. 2018;22(5):1623-1639.
108. Zadeh LA. Is there a need for fuzzy logic? *Inform Sci*. 2008;178(13):2751-2779.
109. Hamamoto AH, Carvalho LF, Sampaio LDH, Abrão T, Proença ML Jr. Network anomaly detection system using genetic algorithm and fuzzy logic. *Expert Syst Appl*. 2018;92:390-402.
110. Vorobiev E, Petrenko S, Kovaleva I, Abrosimov I. Analysis of computer security incidents using fuzzy logic. Proceedings of the 2017 XX IEEE International Conference on Soft Computing and Measurements (SCM); 2017:369–371; IEEE.
111. Alali M, Almogren A, Hassan MM, Rassan IA, Bhuiyan MZA. Improving risk assessment model of cyber security using fuzzy logic inference system. *Comput Secur*. 2018;74:323-339.
112. Joshi C, Ranjan RK, Bharti V. A fuzzy logic based feature engineering approach for botnet detection using ANN. *J King Saud Univ Comput Inf Sci*. 2021;34:6872-6882.
113. Saidi F, Trabelsi Z, Ghazela HB. Fuzzy logic based intrusion detection system as a service for malicious port scanning traffic detection. Proceedings of the 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA); 2019:1–9; IEEE.
114. Haripriya A, Kulothungan K. Secure-MQTT: an efficient fuzzy logic-based approach to detect DoS attack in MQTT protocol for internet of things. *EURASIP J Wirel Commun Netw*. 2019;2019(1):1-15.
115. Stephan G, Pascal H, Andreas A. Knowledge representation and ontologies. In: Studer R, Grimm S, Abecker A, eds. *Semantic Web Services: Concepts, Technologies, and Applications*. Berlin: Springer; 2007:51-105.
116. Maedche A, Staab S. Ontology learning for the semantic web. *IEEE Intell Syst*. 2001;16(2):72-79.
117. Sarker IH, Colman A, Han J. RecencyMiner: mining recency-based personalized behavior from contextual smartphone data. *J Big Data*. 2019;6(1):1-21.
118. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Burlington, MA: Morgan Kaufmann; 1988.
119. Nunes RC, Colomé M, Barcelos FA, Garbin M, Paulus GB, Silva LADL. A case-based reasoning approach for the cybersecurity incident recording and resolution. *Int J Softw Eng Knowl Eng*. 2019;29(11n12):1607-1627.
120. Lansley M, Polatidis N, Kapetanakis S, Amin K, Samakovitis G, Petridis M. Seen the villains: detecting social engineering attacks using case-based reasoning and deep learning. Proceedings of the ICCBR Workshops; 2019:39–48.
121. Al-Mousa MR. Analyzing cyber-attack intention for digital forensics using case-based reasoning. arXiv preprint arXiv:2101.01395, 2021.
122. Kidmose E, Stevanovic M, Pedersen JM. Detection of malicious domains through lexical analysis. Proceedings of the 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security); 2018:1–5; IEEE.
123. Perera I, Hwang J, Bayas K, Dorr B, Wilks Y. Cyberattack prediction through public text analysis and mini-theories. Proceedings of the 2018 IEEE International Conference on Big Data (Big Data); 2018:3001–3010; IEEE.
124. L'Huillier G, Hevia A, Weber R, Rios S. Latent semantic analysis and keyword extraction for phishing classification. Proceedings of the 2010 IEEE international conference on intelligence and security informatics; 2010:129–131; IEEE.
125. Van Hee C, Jacobs G, Emmery C, et al. Automatic detection of cyberbullying in social media text. *PloS One*. 2018;13(10):e0203794.
126. Sekeres J, Ormandjieva O, Suen C, Hamel J. Advanced data preprocessing for detecting cybercrime in text-based online interactions. Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence; 2020:416–424; Springer.
127. Alves F, Bettini A, Ferreira PM, Bessani A. Processing tweets for cybersecurity threat awareness. *Inf Syst*. 2021;95:101586.
128. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. *Comput Intell Neurosci*. 2018;2018:7068349.
129. Rao RS, Ali ST. A computer vision technique to detect phishing attacks. Proceedings of the 2015 5th International Conference on Communication Systems and Network Technologies; 2015:596–601; IEEE.
130. Corum A, Jenkins D, Zheng J. Robust PDF malware detection with image visualization and processing techniques. Proceedings of the 2019 2nd International Conference on Data Intelligence and Security (ICDIS); 2019:108–114; IEEE.
131. Straub J. Cybersecurity considerations for image pattern recognition applications. Proceedings of the 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR); 2018:1–6; IEEE.
132. Vidal JM, Monge MAS, Villalba LJG. A novel pattern recognition system for detecting android malware by analyzing suspicious boot sequences. *Knowl-Based Syst*. 2018;150:198-217.

133. He K, Kim DS. Malware detection with malware images using deep learning techniques. Proceedings of the 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE); 2019:95–102; IEEE.
134. Freitas S, Duggal R, Chau DH. MalNet: a large-scale cybersecurity image database of malicious software. arXiv preprint arXiv:2102.01072, 2021.
135. Kyrkou C, Papachristodoulou A, Kloukiniotis A, et al. Towards artificial-intelligence-based cybersecurity for robustifying automated driving systems against camera sensor attacks. Proceedings of the 2020 IEEE Computer Society Annual Symposium on VLSI (ISVLSI); 2020:476–481; IEEE.
136. Varzaneh ZA, Rafsanjani MK. Intrusion detection system using a new fuzzy rule-based classification system based on genetic algorithm. *Intell Decis Technol*. 2021;15:231-237.
137. Onah JO, Abdulhamid SM, Abdullahi M, Hassan IH, Al-Ghusham A. Genetic algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment. *Mach Learn Appl*. 2021;6:100156.
138. Fatima A, Maurya R, Dutta MK, Burget R, Masek J. Android malware detection using genetic algorithm based optimized feature selection and machine learning. Proceedings of the 2019 42nd International conference on telecommunications and signal processing (TSP); 2019:220–223; IEEE.
139. Biggio B, Fumera G, Roli F. Pattern recognition systems under attack: design issues and research challenges. *Int J Pattern Recognit Artif Intell*. 2014;28(07):1460002.
140. Gardiner J, Nagaraja S. On the security of machine learning in malware C&C detection: a survey. *ACM Comput Surv*. 2016;49(3):1-39.
141. Chen S, Xue M, Fan L, et al. Automated poisoning attacks and defenses in malware detection systems: an adversarial machine learning approach. *Comput Secur*. 2018;73:326-344.
142. Stokes JW, Wang D, Marinescu M, Marino M, Bussone B. Attack and defense of dynamic analysis-based, adversarial neural malware detection models. Proceedings of the MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM); 2018:1–8; IEEE.
143. Kravchik M, Shabtai A. Efficient cyber attack detection in industrial control systems using lightweight neural networks and PCA. *IEEE Trans Dependable Secur Comput*. 2021;19:2179-2197.
144. Alzantot M, Sharma Y, Elgohary A, Ho BJ, Srivastava M, Chang KW. Generating natural language adversarial examples. arXiv preprint arXiv:1804.07998, 2018.
145. Specht F, Otto J, Niggemann O, Hammer B. Generation of adversarial examples to prevent misclassification of deep neural network based condition monitoring systems for cyber-physical production systems. Proceedings of the 2018 IEEE 16th International Conference on Industrial Informatics (INDIN); 2018:760–765; IEEE.
146. Xi B. Adversarial machine learning for cybersecurity and computer vision: current developments and challenges. *WIREs Comput Stat*. 2020;12(5):e1511.
147. Lin WC, Ke SW, Tsai CF. CANN: an intrusion detection system based on combining cluster centers and nearest neighbors. *Knowl-Based Syst*. 2015;78:13-21.
148. Moon D, Im H, Kim I, Park JH. DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks. *J Supercomput*. 2017;73(7):2881-2895.
149. Syed NF, Baig Z, Ibrahim A, Valli C. Denial of service attack detection through machine learning for the IoT. *J Inf Telecommun*. 2020;4(4):482-503.
150. Javed Y, Rajabi N. Multi-layer perceptron artificial neural network based IoT botnet traffic classification. Proceedings of the Future Technologies Conference; 2019:973–984; Springer.
151. Hong T, Choi C, Shin J. CNN-based malicious user detection in social networks. *Concurr Comput Pract Exp*. 2018;30(2):e4163.
152. Xiao X, Zhang D, Hu G, Jiang Y, Xia S. CNN-MHSA: a convolutional neural network and multi-head self-attention combined approach for detecting phishing websites. *Neural Netw*. 2020;125:303-312.
153. Shi WC, Sun HM. DeepBot: a time-based botnet detection with deep learning. *Soft Comput*. 2020;24:16605-16616.
154. Mozzaquatro BA, Agostinho C, Goncalves D, Martins J, Jardim-Goncalves R. An ontology-based cybersecurity framework for the Internet of Things. *Sensors*. 2018;18(9):3053.
155. Arogundade OT, Abayomi-Alli A, Misra S. An ontology-based security risk management model for information systems. *Arab J Sci Eng*. 2020;45(8):6183-6198.
156. Dey AK. Understanding and using context. *Pers Ubiquitous Comput*. 2001;5(1):4-7.
157. Sarker IH, Alqahtani H, Alsolami F, Khan AI, Abushark YB, Siddiqui MK. Context pre-modeling: an empirical analysis for classification based user-centric context-aware predictive modeling. *J Big Data*. 2020;7(1):1-23.
158. Sarker IH. Context-aware rule learning from smartphone data: survey, challenges and future directions. *J Big Data*. 2019;6(1):1-25.

How to cite this article: Sarker IH. Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview. *Security and Privacy*. 2023;6(5):e295. doi: 10.1002/spy2.295