



# Cybersecurity Systems with Automated Artificial Intelligence Applied in Healthcare

**Cheryl Ann Alexander<sup>1</sup> and Lidong Wang<sup>2\*</sup>**

<sup>1</sup>Institute for IT Innovation and Smart Health, Mississippi, USA

<sup>2</sup>Institute for Systems Engineering Research, Mississippi State University, Mississippi, USA

**\*Corresponding author:** Lidong Wang, Institute for Systems Engineering Research, Mississippi State University, Mississippi, USA

**Received Date:** April 22, 2024

**Published Date:** May 21, 2025

## Abstract

Artificial intelligence (AI)/machine learning (ML) can increase the speed and scale of cybersecurity defenses. Cybersecurity automation can be enhanced using AI/ML. However, threat actors can also use AI and often change and improve their attack strategies. Some malicious attacks are on vulnerable AI models. This paper deals with cybersecurity systems with automated artificial intelligence (AI). Several important topics are introduced or discussed. They include frameworks and practices of automated cyberattacks and cybersecurity automation; behavioral differences among threat actors; an AI asset taxonomy system and asset categories; threat/risk categorization, threat-actor tactics, and TTPs (tactics, techniques, and procedures); and the cyber kill chain and offensive AI capabilities. Patient privacy concerns and data breaches are the priority for healthcare cybersecurity. Automated AI cybersecurity systems in healthcare are presented as a case study in this paper.

**Keywords:** Cybersecurity; Automation; Information; Artificial Intelligence (AI); Machine Learning (ML); Offensive AI; Healthcare

## Introduction

Systems that utilize artificial intelligence (AI)/machine learning (ML) to handle cybersecurity problems could better decide adversarial intent(s) and future actions by connecting sequences of actions to threat actor intent(s) [1]. AI-powered cybersecurity tools learn and adapt to new cyber threats despite the progressively more sophisticated techniques cyber threat actors employ. AI evolves to detect and deflect these adapted methods. AI can correlate data from various sources, giving cybersecurity professionals a broad view of possible threats. Early detection of advanced, multi-stage threat attacks that might otherwise go undetected is possible.

AI/ML can eliminate some threats (e.g., phishing). AI/ML can make system owners detect earlier and deter some of the distributed denial-of-service (Dos) attacks more efficiently and prevent data leakage and network penetration. AI/ML can dramatically increase the speed and scale of cybersecurity defenses by using advanced tools to increase automation. Robust cybersecurity requires automation [2]. A comprehensive and lightweight hybrid feature selection and an ensemble classifier were proposed that selected a few relevant features and provided an accurate and consistent classification of most attacks [3].

Table 1 [4] lists the advantages and disadvantages of some AI techniques used in the intrusion detection system (IDS). A neural network (NN) mimics the human brain to establish an information processing system with interconnected nodes working with each other to solve a problem. Artificial immune system (AIS) is inspired by the biological immune system which transforms a biological

model and the functions of the immune system into a mathematical model to facilitate problem-solving. The genetic algorithm (GA) is an optimization method to find approximate solutions to search problems. Fuzzy logic (FL) deals with reasoning that is approximate rather than exact or fixed.

**Table 1:** Advantages and Disadvantages of the AI-based IDS.

Techniques	Advantages	Disadvantages
NN-based	<ul style="list-style-type: none"> <li>Multi-layers in NN increase the classification efficiency.</li> <li>Effectively classifies unstructured network packets.</li> </ul>	<ul style="list-style-type: none"> <li>Requires many samples for effective training.</li> <li>Requires long time for training.</li> <li>Less flexible than other techniques.</li> </ul>
AIS-based	<ul style="list-style-type: none"> <li>Good detection accuracy</li> </ul>	<ul style="list-style-type: none"> <li>Many parameters</li> </ul>
GA-based	<ul style="list-style-type: none"> <li>Good efficiency</li> <li>Selecting the best features for detection.</li> </ul>	<ul style="list-style-type: none"> <li>A complicated method to represent a problem.</li> <li>Used in a specific manner rather than a general manner.</li> </ul>
FL-based	<ul style="list-style-type: none"> <li>Good flexibility to some uncertain problems.</li> </ul>	<ul style="list-style-type: none"> <li>Lower detection accuracy than NN.</li> </ul>

Intelligence in cybersecurity is frequently referred to as computational intelligence, and it characterizes a system’s capability to extract knowledge from cybersecurity data to conclude or through experimental observation, as well as from learning specific cybersecurity task(s). Intelligent systems could help with various cybersecurity problems [5]. There are the following types of AI methods [5]:

- Analytical AI—extracts patterns or insights from data to offer suggestions.
- Functional AI—executes an action based on extracted knowledge or insights rather than providing suggestions.
- Interactive AI—automates communication without sacrificing interactions.
- Visual AI—using computer vision and extracting insights

from visual data or images.

- Textual AI—using text analytics and mining as well as language processing.
- Hybrid AI—combining the AI as mentioned above.

Swarm intelligence/swarm optimization is a complicated ML method that uses the collective intelligence of a group of cooperative agents to solve problems (e.g. intrusion detection) through evolutionary computations [6]. Table 2 [7] lists the ML types with tasks and model-building methods. ML has been used in cybersecurity and threat intelligence. It helps find insider threats, detect malware in encrypted traffic, secure data in the cloud by uncovering suspicious activity, etc. For example, clustering has been employed to recognize cyber anomalies, policy violations, etc. Deep learning (DL) has been utilized in data analytics on large security datasets [8].

**Table 2:** Types of ML.

Learning Types	Tasks	Model Building
Unsupervised	Associations, clustering, dimension reduction	Models or algorithms learn from unlabeled data (data-driven method)
Supervised	Classification, regression	Models or algorithms learn from labeled data (task-driven method)
Semi-supervised	Classification, clustering	Models are created using combined data (unlabeled + labeled)
Reinforcement	Classification, control	Models are created based on reward or penalty (environment-driven method)

The main purpose of research in this paper is to deal with cybersecurity systems with automated AI. The remainder of this paper will be organized as follows: the second section introduces frameworks and practices of automated cyberattacks and cybersecurity automation; the third section presents behavioral differences among threat actors; the fourth section introduces an AI asset taxonomy system and asset categories; the fifth section deals with threat/risk categorization, threat-actor tactics, and TTPs; the sixth

section introduces the cyber kill chain and offensive AI capabilities; the seventh section presents automated AI cybersecurity systems in healthcare; and the eighth section is the conclusion.

### Frameworks and Practices of Automated Cyberattacks and Cybersecurity Automation

An automated approach to strengthening security against attacks is essential. A preemptive prediction-based automated cy-

berattack framework was proposed that reveals four main aspects and is shown in Figure 1 [9]. First is the structured/nonstructured data collection and the refining of the collected data for attack simulations. Second is the vulnerability-based knowledgebase that collects information on attacks by third parties, including viruses and attack patterns, vulnerabilities, and technologies against the attacks. Third is the attacker-oriented attack strategic simulation. Fourth is the attack prediction with the information from the AI attack simulation.

An ontology-based cybersecurity framework was proposed that extends well-known cybersecurity ontologies to specifically model and manage threats imposed on systems, services, and applications that rely on AI (Preuveneers and Joosen, 2024). Cybersecurity automation decreases human dependency while providing a better approach to detecting, responding to, protecting against, and recovering from cyberattacks. Cybersecurity can be automated with advanced methods such as AI/ML [5].

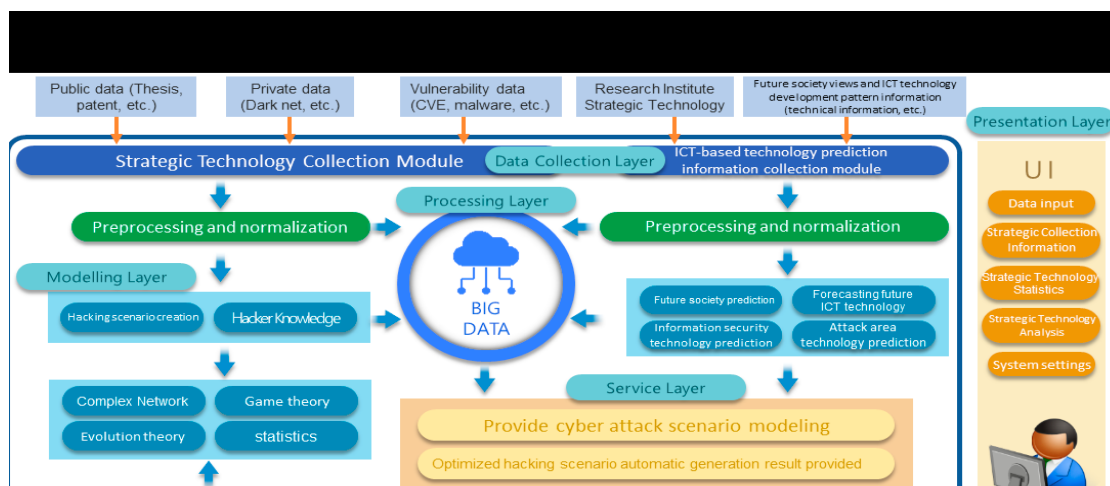


Figure 1: A proposed preemptive prediction-based automated cyberattack framework.

## Behavioral Differences Among Threat Actors

Specific to AI cybersecurity systems, there are behavioral differences among the common groups of threat actors, including cybercriminals, insiders, terrorists, hacktivists, script kiddies, competitors, and nation-state actors. Cybercriminals use multiple sources of massive data gathered and analyzed by AI to collect, propose, target, and devise false identities of potential victims in convincing and efficient social media attacks. Each organization must acknowledge and seriously contemplate the possible harm caused by insiders who misuse AI for personal wealth or just to even the score. Insider threats commit actions for various reasons; psychological beliefs or other personal motivations are major reasons [10].

Terrorist groups use cyberspace to initiate cyberattacks that use complex programs and the Internet. Terrorists use the web for collecting information, recruitment, planning, coordinating, and financing malicious attacks. Unfortunately, the malicious abuse and use of AI by terrorists is widespread. Hackers/hacktivists use ML vulnerabilities to manipulate the ML system's integrity, confidentiality, and availability. Hackers use multiple ways to leverage AI to make their hacks more successful: data breaches and data mining, phishing, stealing passwords, evading malware detection, etc. Script kiddies use AI to develop and launch malicious attacks. Script kiddies use AI for automating tasks such as exploitation and reconnaissance, evading detection by security software, etc. They

also construct targeted malware attacks tailored to specific victims. Competitors scan for vulnerabilities in their network servers before they attack rivals. If they find any vulnerabilities, they report and predict when a cyberattack may happen. Competitors simultaneously create their network defenses to protect their servers. Between 2011 and 2014, a nation-state actor simply called APT28, targeted mainly military, embassy, and defense contractor personnel from the USA and its allies. The goal was to perform persistent political and economic espionage on the targets. A behavioral model of APT28 to exfiltrate all email messages from the targeted users' mailboxes, was presented [1].

Both Microsoft and OpenAI have warned that nation-state actors have used ChatGPT to automate several phases in their attack chains, which include social media attacks and reconnaissance. The rise of AI-augmented cyber defenses incorporated into national defense postures will probably be vulnerable to 'poisoning' attacks that predict, manipulate, and subvert the functionality of defensive algorithms. Input attacks are a kind of contestation that seeks to essentially mislead an AI system and skew its efforts to classify activity patterns. In contrast to input attacks, the 'poisoning' attacks essentially seek to compromise the AI programming, which is utilized in enemy systems [11]. AI also makes counterintelligence efforts more powerful. Counterintelligence leverages all national security authorities and resources available to fight hostile nations seeking to damage the USA [10].

## An AI Asset Taxonomy System and Asset Categories

Using asset taxonomies to tag and organize assets is a powerful way to streamline consistency in the asset library. Figure 2 shows a created block diagram of an AI asset taxonomy system. Data comes in various forms such as visual data (made up of images); voice data; textual data; and numerical data. Examples of numerical data are accuracy, true positive (TP), true negative (TN), false positive (FP), false negative (FN), precision, recall, false acceptance rate (FAR), false rejection rate (FRR), and F1 score. Natural language processing (NLP) is an AI technology for textual and voice data. Computer vision is an AI technology for visual data. The definition of an AI model is a program, algorithm, or tool based on specific datasets that can recognize patterns to arrive at a decision. For example, an ML model trains a dataset and utilizes a mathematical formula to predict new data or future events. Regression models and classification models are examples of ML models. Another ML model is a DL model or a neural network trained for learning to perform a task. The most popular AI models include the following:

- Linear regression
- Deep neural networks
- Logistic regression
- Decision trees
- Linear discriminant analysis
- Naive Bayes
- Support vector machines
- Learning vector quantization
- k-nearest neighbors
- Random forest

- Linear regression

There are several types of AI actors. Third-party entities are providers, vendors, developers, and data evaluators. These third-party entities design, develop, and deploy AI but are external to the organization. Individuals or groups that utilize an AI system for precise reasons are end users. The public will have positive and negative impacts directly related to AI technology. The public includes individuals, consumers, and communities most associated with the context surrounding the development or deployment of the AI system. Other AI actors guide for managing AI risks including standards-developing organizations, environmental groups, researchers, etc. As for processes in AI, AI systems work using algorithms and data. Massive amounts of data are collected and applied to mathematical models and algorithms to identify patterns and make predictions. AI tools and environment include the following:

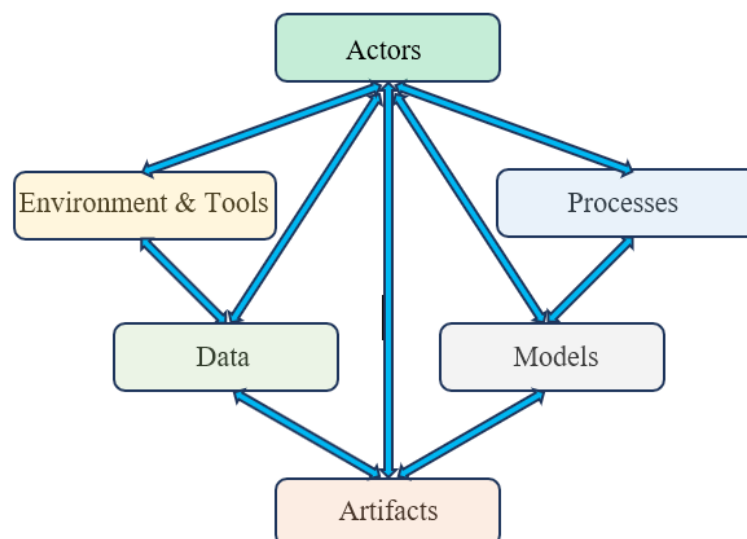
- IDE (Integrated Development Environment) tools for AI
  - Visual Studio Code, PyCharm, and Eclipse.
- ML tools and AI frameworks
  - TensorFlow, Keras, sci-kit-learn, and PyTorch.
- Learning resources and communities

There are some of the best resources for real intelligence on AI data structures and algorithms, such as

- Coursera and DataCampC

There are some of the top software developer communities on the interwebs, such as

- GitHub, Stack Overflow, and Reddit



**Figure 2:** Block diagram of an AI asset taxonomy system.

All digital products used in an AI Tool are described in AI artifacts. This can include input, output, or intermediate results processed by tools. There are six primary types of artifacts: data, algorithm, benchmark, application, model, and knowledge. The most common ML artifacts are features, interference data, models, and training data.

### Threat/Risk Categorization, Threat-actor Tactics, and TTPs

AI systems can be categorized into four levels using a risk-based approach: minimal or no risk, limited risk, high risk, and unacceptable risk. Cybersecurity experts must apply appropriate risk management principles to AI development. Data leakage, data poisoning, and data integrity attacks can occur at any stage of the AI development or supply chain. Consumer privacy, danger to humans, biased programming, and unclear legal regulation are a few of the biggest risks. High-risk AI systems include safety components in the management or operation of critical digital in-

frastructure, water supply, heating, road traffic, gas, and electricity. Threat-actor tactics are the high-level description of the behaviors and strategies (especially new threat-actor offense strategies) of a malicious actor. A tactic includes a set of actions and behaviors that can be employed by the threat actor to achieve specific objectives.

Tactics will need to evolve to prepare for future cybersecurity. Techniques are non-specific guidelines and intermediate methods to describe how a tactic action can be realized. Procedures refer to the sequence of actions that are performed using a technique to execute an attack tactic. Malware development and deployment (trojans, viruses, ransomware, etc.); exploitation of vulnerable software; phishing attacks and social engineering to gain access to passwords; and shoulder surfing-the practice of looking “over the shoulder” of someone to take a photo of valuable information or jot down information are examples of tactics used by threat actors. A model regarding tactics, techniques, and procedures (TTPs) and cyberspace was developed, which is shown in Figure 3 [1].

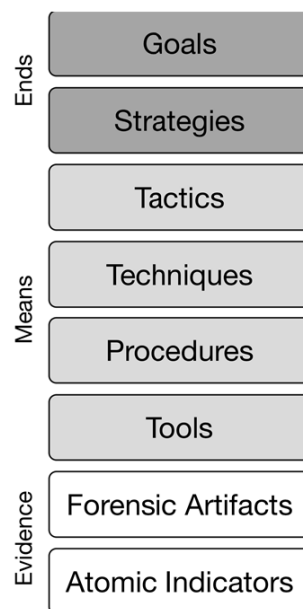


Figure 3: A model regarding TTPs.

Figure 4 [1] shows how ML fits into a data pipeline used to inform a human controller of a possible attack. The cognitive agent has prior knowledge of previously seen attack models and an in-

ventory of assets. The human controller could inform the cognitive agent whether it was correct or not to facilitate reinforcement learning.

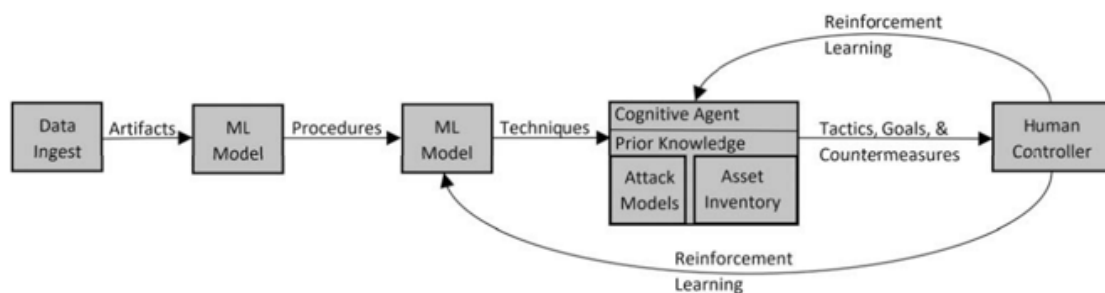


Figure 4: An ML pipeline for detecting TTPs.

### The Cyber Kill Chain and Offensive AI Capabilities

A developed concept of the cyber kill chain is shown in Table 3 [12,9]; however, experts suggested that a strengthened defense be

required for the internal firewall and that reconnaissance on the internal and an effective recovery procedure on the weaponization stage also be required [12].

**Table3:** The cyber kill chain defense strategy.

Steps	Description
Reconnaissance	An intruder selects a target, investigates it, and tries to detect vulnerabilities in the target network.
Weaponization	The intruder creates a remote access malware weapon (e.g., virus or worm) that is tailored to one or more vulnerabilities.
Delivery	The intruder transmits the weapon to the target.
Exploitation	The program code of the malware weapon triggers, which acts on the target network to exploit one or more vulnerabilities.
Installation	The malware weapon installs an access point that is usable for the intruder.
Command & control	The malware enables the intruder to have “hands-on-the-keyboard” persistent access to the target network.
Actions on objectives	The intruder acts to achieve the goal(s) such as encryption for a ransom, data destruction, or data exfiltration.

Threat actors frequently change and improve their attack strategies using advanced technologies such as AI, which is called AI-based cyberattacks. AI enables the automation of cyber defense tasks, for example, vulnerability evaluation, threat intelligence processing, intrusion detection, and incident response [13]. Figure 5 [13] illustrates threat actors’ activities within the cyberattack life cycle and offensive AI capabilities (allowing an attacker to conduct an attack on a large scale). The activities include the following stages:

1) Reconnaissance (AI can be used for observing normal behav-

iors and operations),

2) Weaponization,

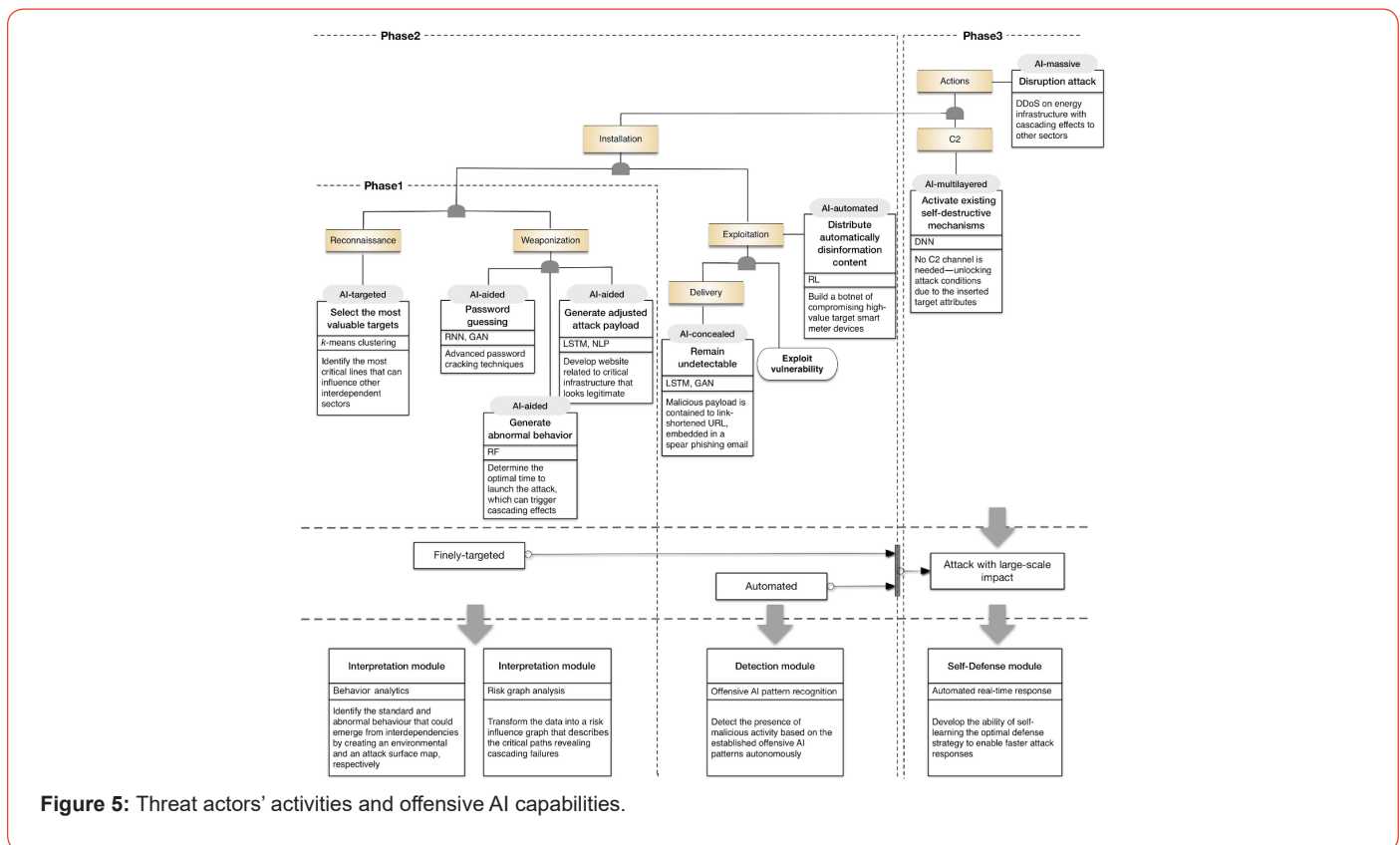
3) Delivery,

4) Exploitation,

5) Installation,

6) Command and control (no C2 channel are needed due to using AI), and

7) Actions on objectives.



**Figure 5:** Threat actors’ activities and offensive AI capabilities.

## Automated AI Cybersecurity Systems in Healthcare

Data security and patient privacy are essential to a robust cybersecurity plan for healthcare facilities. When considering adopting AI medical documentation solutions, the most critical factor is the security of patient data. Healthcare organizations must put robust patient data security measures, vigorous encryption methods, and tough data governance policies into AI documentation solutions that process and store crucial patient data according to standard HIPAA (the Health Insurance Portability and Accountability Act) compliance [14]. Therefore, certain controls should be in place to record and monitor data access, ensuring authorized staff are the only individuals able to view or modify data. Audit trails, access controls, and user authentication ensure that only authorized personnel can access the patient data. However, AI has many sound uses for healthcare providers. For example, providers and radiologists use AI to identify patterns and diagnose neurological diseases, cardiovascular disorders, cancer, etc. Although the transition from paper to a digital environment in healthcare has brought numerous beneficial changes, many cybersecurity challenges have also been introduced, such as:

- Human error and insider threats—for healthcare employees, the simple act of accidentally sending an email with patient data to the wrong email address can result in a significant breach. Malicious threat actors can then exploit this access for personal gain. For example, threat actors sell patient data to identity thieves.
- Interconnected systems and third-party risks—it is necessary to question the security of the Internet of Things (IoT). Threat actors can remotely manipulate medical devices, potentially leading to errors in dosage, rate, rhythm, etc., depending on the type of device. For example, a patient could receive the wrong dose of insulin if an insulin pump is manipulated, resulting in injury or death. Patient safety and risk should be seriously evaluated.
- Patient privacy violations and data breaches—sensitive data such as names, social security numbers, addresses, etc. are exposed by a data breach.
- Security frameworks and regulatory compliance—HIPAA and GDPR (General Data Protection Regulation) impose stringent standards on patient data protection.

Although AI technology enhances healthcare outcomes, several cybersecurity risks must be addressed. Patient privacy concerns and data breaches are the priority for healthcare cybersecurity professionals. Malicious attacks on vulnerable AI models, ransomware attacks, and supply chain vulnerabilities can also interfere with patient care. AI application security and infrastructure include hiding model parameters to protect against model attacks, integrity validations, oversight of AI model behavior, minimizing privileges of AI models, data quality assurance and integrity validations, and monitoring and incident detection (detects abuse) must also be addressed by cybersecurity professionals. Healthcare entities must have a strategic vision for vigorous cybersecurity, which is described as follows:

- Cybersecurity measures should be strengthened for healthcare entities. This includes encryption and secure transmission of patient data, and a tough authentication/access control process.
- Ethical standards must also be implemented specifically for AI development and deployment so that ethical guidelines govern healthcare frameworks and the promotion of interdisciplinary collaboration and stakeholder engagement.
- The IT cybersecurity team is required to continually monitor and update the system with regularly scheduled system audits and vulnerability assessments. Furthermore, regular, and continuous training and cybersecurity awareness programs for healthcare professionals are crucial.

Protecting patient data from breaches and compliance with HIPAA and other privacy regulations is a key concern for professionals implementing AI in healthcare. Threat actors can also target AI models. Antagonistic cyberattacks can control AI algorithms to give incorrect diagnostic or treatment recommendations, threatening patient lives. Data poisoning attacks can also manipulate training data (or labels of the data) and the behavior of the AI model leading to sabotage of the model or encouraging decisions in favor of the threat actor. In the healthcare environment, implementing specialized security measures is essential to a healthy and robust cybersecurity system. There are the following specific measures:

- A multi-layered, multi-point defense is a cybersecurity strategy that incorporates defensive AI strategies and includes multiple elements such as firewalls, advanced threat detection, intrusion detection systems, etc.
- Access control and robust data encryption also protect sensitive patient data and restrict access so that only authorized personnel can access the data. Strong access controls prevent unauthorized access to private patient records, inappropriate access to the AI systems, underlying training modules, infrastructure, etc.
- Third-party vendor assessments are necessary to evaluate the security systems of third-party vendors or contractors. This includes vendors that use AI, provide AI systems, or use AI systems in their security systems. This ensures that third parties adhere to strict security standards and protocols.
- AI should also be used to construct a healthy incident response plan. Healthcare entities also need regular security audits and updates for their AI systems and the overall infrastructure. Healthcare facilities use patch management to ensure that AI hardware and software components are up-to-date and protected with the latest security updates and patches.
- In the healthcare environment, most employees are not trained in cybersecurity practices and policies. Therefore, an ongoing, strong, solid training and awareness program is necessary for all healthcare staff.

Current networks generate a high volume of alerts. Analysts must spend the time and use expertise to investigate and respond to numerous high-priority incidents. AI reduces this burden by au-

tomating manageable security operations to increase efficiency and ensure manageable security operations. AI additionally enhances threat intelligence capacities through automation. AI automates the collecting, analyzing, and sharing of security-related data with other healthcare facilities to identify evolving threats, trends, and other markers of malicious threat actors without human involvement. AI systems analyze and correlate data in real time. Security teams can then identify actionable events and contexts during a security incident. When cybersecurity professionals can rapidly recognize the scope and severity of an attack, they can initiate a prompt response and necessary containment measures, minimizing the impact of a breach. Cybersecurity specialists can also use AI in patching and vulnerability management.

## Conclusion

Cybersecurity can be automated with AI/ML. Robust cybersecurity requires automation. AI can also enhance threat intelligence capacities through automation. Threat actors frequently change and improve their attack strategies using AI, which is called AI-based cyberattacks. AI can correlate data from various sources. AI also helps analyze and correlate data in real time. Although AI technology enhances healthcare outcomes, several cybersecurity risks must be addressed. Patient privacy concerns and data breaches are the priority for healthcare cybersecurity professionals. Protecting patient data from breaches and compliance with HIPAA and other privacy regulations is a key concern for professionals implementing AI in healthcare. Malicious attacks on vulnerable AI models, ransomware attacks, and supply chain vulnerabilities can also interfere with patient care. AI-based cyberattacks and offensive AI capabilities are important future research topics.

## Acknowledgements

The authors would like to express thanks to Technology and Healthcare Solutions, USA for its help.

## Conflict of interest

The authors would like to announce that there is no conflict of interest.

## Ethics

In this article, ethical principles related to scientific research articles are observed. The corresponding author confirms that both authors have read, revised, and approved the paper.

## Consent for publication

Not applicable.

## Availability of data

Not applicable.

## References

1. Maymí F, Bixler R, Jones, R, Lathrop S (2017) Towards a definition of cyberspace tactics, techniques and procedures. In *2017 IEEE international conference on big data (big data)* pp. 4674-4679.
2. Bresniker K, Gavrilovska A, Holt J, Milojicic D, Tran T (2019) Grand challenge: Applying artificial intelligence and machine learning to cybersecurity. *Computer* 52(12): 45-52.
3. Jaw E, Wang X (2021) Feature selection and ensemble-based intrusion detection system: an efficient and comprehensive approach. *Symmetry* 13(10): 1764.
4. Alrajhi AM (2020) A survey of Artificial Intelligence techniques for cybersecurity improvement. *Int. J. Cyber-Secur. Digit. Forensic* 9: 34-41.
5. Sarker IH (2023) Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview. *Security and Privacy* 6(5): e295.
6. Gyamfi NK, Goranin N, Ceponis D, Čenys HA (2023) Automated system-level malware detection using machine learning: A comprehensive review. *Applied Sciences* 13(21): 11908.
7. Sarker IH (2022) AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science* 3(2): 158.
8. Sarker IH (2021) Machine learning: Algorithms, real-world applications and research directions. *SN computer science* 2(3): 160.
9. Ryu S, Kim J, Park N, Seo Y (2021) Preemptive Prediction-Based Automated Cyberattack Framework Modeling. *Symmetry* 13(5): 793.
10. Ali S (2019) Cybersecurity support of insider threat operations: DoD regulation and constitutional compliance. *George Mason University Civil Rights Law Journal* 30:1-64.
11. Whyte C (2020) Poison, persistence, and cascade effects. *Strategic Studies Quarterly* 14(4): 18-46.
12. Park N, Kang N (2015) Mutual authentication scheme in secure Internet of things technology for comfortable lifestyle. *Sensors*, 16(1): 20.
13. Kaloudi N, Li J (2020) The AI-based cyber threat landscape: A survey. *ACM Computing Surveys (CSUR)*, 53(1): 1-34.
14. Preuveneers D, Joosen W (2024) An Ontology-Based Cybersecurity Framework for AI-Enabled Systems and Applications. *Future Internet* 16(3): 69.