Artificial Intelligence in Cybersecurity: A Comprehensive Review of Techniques, Applications, and Ethical Challenges

J. Ahamed Aathil, Nivitha Rajendran, M.R.M. Hanan, and U.A.F. Munaffara

Abstract This paper examines how artificial intelligence is transforming the evolving landscape of cybersecurity. It reviews key applications of AI in threat detection, incident response, security analytics, and malware analysis. In particular, the study compares the effectiveness of deep learning techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Artificial Neural Networks (ANNs) in enhancing security system performance. This study follows a systematic literature review (SLR) methodology, analyzing 50 peer-reviewed studies published between 2017 and 2025, retrieved from IEEE Xplore, Springer, ScienceDirect, ACM Digital Library, and Academia. Overall, these AI-based approaches demonstrate significant potential compared to traditional systems for processing large datasets, identifying patterns, and detecting anomalies associated with malicious activities. The paper also discusses the automated auditing of smart contracts and vulnerability assessment using artificial intelligence and natural language processing techniques. Furthermore, it examines ethical challenges associated with AI adoption in cybersecurity, including transparency, accountability, and bias-free algorithm design and deployment. Finally, the study highlights emerging research and development trends, such as the application of AI in counterterrorism cybersecurity and AI-generated simulated secure environments, which reflect evolving patterns in modern cybersecurity practices.

***Index Terms*— Artificial Intelligence in Security, AI-driven Security Systems, Machine Learning for Security, Cybersecurity and Artificial Intelligence, AI-based**

I. Introduction

INTERNET and e-technologies have grown at an extremely high rate contributing to the increased global connectivity, yet exposing the world to the advanced cyber threats [1], [2]. Consequently, the protection of confidential data and creation of confidence in digital infrastructure have been of high priority to governments, organizations and individuals [3]. Conventional cybersecurity solutions that are reactive, static, and signature-based have a hard time keeping up with the changing strategies employed by current cyber attackers. The level of sophistication used by cybercriminals today can overcome traditional defense systems and cybercriminals have to be dealt with by more intelligent and dynamic security measures. In this regard, artificial intelligence (AI) has become a game changer as a facilitator, and it has essentially changed all defensive approaches to issues that involve the cybersecurity environment [4]. Security systems using machine learning (ML) models are AI systems that utilize learning algorithms to process lots of data, identify abnormal behavior, and subtle signs of compromise that conventional systems cannot effectively identify [5]. Artificial neural networks (ANN) and support vector machines (SVM), are just some of the techniques that have shown great success in the detection of anomalies that could be used to better detect intrusions through more efficient and effective intrusion detection system (IDS) [6]. Therefore, AI enhanced IDS models tend to be more accurate, precise and recollective than traditional security systems [7].

Deep learning as a subdivision of AI empowers threat detection further by training models on large amounts of malicious behavior and, therefore, generalizing to novel and otherwise unknown attack vectors, especially zero-day exploits [8]. Moreover, the AI boosts incident response by detecting, classifying, and prioritizing threats automatically, thus reducing the response time and minimizing possible system damage [9]. Artificial intelligence based analytics also support the detection of malware, the prevention of phishing, and vulnerability testing to help an organization transition to proactive mitigation of cyber risks rather than reactive defense [10]. The application of AI in the field of cybersecurity is not limited to traditional understanding of this concept, and the new systems of blockchain and smart contracts become the new fields where AI can assist in the performance of security tasks, identifying fraudulent activity and guaranteeing the integrity of the

J. Ahamed Aathil is a demonstrator at the Department of ICT, South Eastern University of Sri Lanka, Sri Lanka. (Email: ahamedaathil.5@gmail.com)

Nivitha Rajendran is a demonstrator at the Department of ICT, South Eastern University of Sri Lanka, Sri Lanka. (Email: nivitharajendran98@gmail.com)

M.R.M. Hanan is a demonstrator at the Department of ICT, South Eastern University of Sri Lanka, Sri Lanka. (Email: hananronaldo17@gmail.com)

U.A.F. Munaffara is a is a demonstrator at the Department of ICT, South Eastern University of Sri Lanka, Sri Lanka. (Email: munaffaraumarali@gmail.com)

ISSN: 2961-5992

decentralized system [11]. Nonetheless, as AI enhances security in cyberspace, it also brings in the ethical and technical issues such as challenges in explainability, risk to data privacy, and vulnerability to adversarial attacks on AI models itself. To make the maximum use of AI advantage without creating new weak points, it is necessary to work on these issues. In this regard, the given paper will conduct a thorough review of the various uses of AI in cybersecurity. It investigates the ways AI can be used to improve threat detection, enhance incident response, analyze malware and phishing more thoroughly and the ways it can be used to secure emerging technologies, and evaluates critically whether ethical considerations are needed as AI increasingly becomes integrated into the world in terms of global security frameworks [12].

A. Scope of the Review

This paper focuses on AI-driven cybersecurity techniques applied at the software, network, and system levels, with particular emphasis on machine learning (ML), deep learning (DL), and explainable AI (XAI) approaches. The scope includes applications such as intrusion detection systems, malware analysis, threat intelligence, incident response, IoT security, enterprise security operations, and national security use cases.

The review does not focus on low-level hardware security designs, cryptographic algorithm design, or purely legal/regulatory analyses of artificial intelligence, except where such aspects directly impact AI-enabled cybersecurity systems. By limiting the scope in this manner, the paper concentrates on practical, data-driven, and deployable AI techniques relevant to modern cyber defense environments.

II.Literature Review

A. Introduction

The uncontrolled spread of internet usage, cloud computing, the Internet of Things (IoT), and other digital technologies have created an extremely interconnected world that offers a great deal of convenience and efficiency but, as well, subjects individuals, organizations, and nations to a growing range of cyber threats. The conventional cybersecurity tools such as firewalls, antiviruses software and signature-based intrusion detection systems are reactive in nature and have become unsuitable to counter the advanced strategies used by present day cybercriminals [6]. Cyberattacks have become more sophisticated, common and costly, leading to the need to make proactive, intelligent and dynamic security measures, which can accommodate the dynamism of the threats [11]. Predictive analytics, massive data processing and highly sophisticated automation emblemized as artificial intelligence (AI) has become a game changer in the cybersecurity practice [16]. AI supports detection and prediction of cyber threats and mitigation of cyber-attacks at a level that is impossible with traditional systems due to the possibilities of machine learning, deep learning, and intelligent decision-making models. The following literature review thus discusses the widespread use, performance, and drawbacks of AI in cybersecurity, in sectors like intrusion identification, malware inspection, threat forecasting, automated response to incidents, and ethical frameworks required to maintain the responsible use of AI [48].

B. AI in Threat Detection and Predictive Security

Threat detection represents a state of the art of cybersecurity systems nowadays, and it is based on machine learning (ML) and deep learning (DL) algorithms and processing on large volumes of network traffic, system logs, endpoint actions, and user behavior in real-time [8]. In contrast to the traditional rule-based models, which depend on a set of established signatures, AI models can receive behavioral patterns, discover the slightest abnormalities, and retrieve advanced attacks that would otherwise remain unnoticed by the traditional models. A supervised learning model like Support Vector Machines (SVM), Random Forest, and Artificial Neural Networks (ANN) continues to be very effective in distinguishing between malicious events and legitimate ones since they are able to derive discriminatory actions to structured data set [17], [19]. Applications in these models include intrusion detection, spam filtering, and classifying the malware with great accuracy and flexibility according to various security set-ups.

Deep learning techniques also improve the threat detection performance as they operate on high-dimensional and unstructured security data. Convolutional Neural Networks (CNNs) are promoted to make predictions of spatial-based representations that are derived using network flow patterns, malware binaries, and system call graphs, which make it possible to design sophisticated anomaly detection and behavior-based threat management. Long Short-Term Memory (LSTM) networks and recurrent Neural Networks (RNNs) are especially useful in sequence-based models, which in turn can be useful in identifying the patterns in the logs, attacks in multiple stages, and temporal outliers. Alongside that, Generative Adversarial Networks (GANs) are also currently under investigation to create sample attacks and pose adversarial examples to increase model robustness and train AI systems against previously unseen or zero-day attacks [13].

With constant learning on large and continuously changing data, the system of threat detection based on AI considerably outsmarts conventional security methods in speed, accuracy, scalability, and adaptability of detection [6], [9]. Not only do the systems decrease false positives, but also give rise to predictive security where there is a forecast of possible threats that can be predicted even before it comes into reality. The AI-based threat detection becomes increasingly important as cyberattacks become more sophisticated and automated, and this is one of the aspects of the overall cybersecurity posture that improves its ability to respond, rather than react to the risk.

C. Intrusion Detection Systems (IDS)

The Intrusion Detection Systems (IDS) based on artificial intelligence have become a key use of this technology in cybersecurity to act as an adaptive and intelligent system to identify, stop, and react to cyber threats [21], [23]. In comparison to the traditional IDS solutions, which mainly use detection methods to analyze the network traffic, system logs, and user behaviors with the focus on the signature-based techniques, AI-driven IDS can be used to analyze the network traffic, system logs, and user behavior dynamically and determine the presence of anomaly, suspicious activities, and patterns of potential attacks [6], [13]. The support vector machineries (SVM), the random forests,

ISSN: 2961-5992

the decision trees and the artificial neural networks (ANN) are machine learning (ML) models that are normally used to categorize the network event as a normal or malicious event enhancing the ability to detect events and minimizing false positives [18], [20].

Publicly available benchmark datasets, including KDD'99, NSL-KDD, UNSW-NB15, and CICIDS2017, are often used in order to present these models and provide them with training and validation. The datasets contain labeled examples of normal and attack traffic, which represent a large selection of types of intrusions, including Denial-of-Service (DoS), probing, Remote-to-Local (R2L), and User-to-Root (U2R) attacks. The recent works show that ensemble learning, hybrid ML models and feature selection methods considerable increase the performance of I detectors combining different classifiers, increasing detection rates, and alleviating over-fitting [8], [20].

Moreover, AI-driven IDS systems can adapt to the changing and hitherto unknown lines of attack in real-time, overcoming the shortcomings of more traditional defense tools [13]. Training models such as the Convolutional Neural Networks (CNN) and the Recurrent Neural Networks (RNN) have been progressively incorporated in the IDS platforms to support sequential and high-dimensional data processing to identify sophisticated, multi-phase attacks and zero-day vulnerabilities [23]. Through learning continuously by monitoring new traffic patterns and applying predictive analytics, AI-IDS does not only make intrusion detection more accurate and responsive but also acquires a base layer on which proactive cybersecurity can be implemented on the current digital infrastructure.

D.Malware Analysis and Threat Classification

Artificial intelligence has now become an essential arm in malware detection, classification and mitigation and has become a tool that is far more powerful than traditional signature methodologies. AI-based malware detectors use both static and dynamic features to detect potentially malicious software including file signatures, hash, structural property, runtime behavior, API calls, and system interactions to identify the malicious code even previously unknown or polymorphic code. This two-facet technique opens the way to machine learning (ML) algorithms, including Support Vector Machines (SVM), Random Forests and Artificial Neural Networks (ANN), which can identify files at an exceptionally high level of accuracy, thereby separating between malware and non-malicious software [19]. Malware detection has further been improved by deep learning, especially Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) models, which detect malware in binary malware binaries and logs of malware behavior due to complex temporal and spatial features. LSTM-based models, such as those, are very effective at sequential behavior modeling and allow identifying incubating malware via predictive models, whereas CNN-based systems can identify hierarchical features in binary files and memory dumps to identify subtle malicious behavior signs [12]. AI will also make it possible to organize mass, automatic, malware scanning of both enterprise networks and cloud infrastructures with the ability to classify quickly the type of attack and forecast the trends in the spread of malware [21]. These types of predictive analytics can help organizations have a better understanding of how to focus on remediation, project spread of attacks, and to prevent the damage that may result before it gets out of control. In addition, it has been demonstrated that AI-based malware detection implemented on the Security Information and Event Management (SIEM) and Endpoint Detection and Response (EDR) systems has effectively helped ease the human resource, speed up detection, and enable more prompt response to an incident [41].

E. Adversarial AI and Security Vulnerabilities

Although AI also provides substantial innovations in cybersecurity, it is not free of attacks on weaknesses and misuse. Adversarial attacks are one of the main issues, and the attacker may intentionally alter the input data to defeat the AI models; this may cause misclassification, false negatives, or avoidance of security measures. These attacks are able to take advantage of a training dataset weakness, such as a lack of diversity, an uneven number of classes in a dataset, or biases in the algorithm, or even an opaque model that makes decisions [22]. The sophistication of adversarial methods is a significant risk, especially in high-stakes settings (financial system, critical infrastructure and national security applications). In an attempt to mitigate such challenges, scholars have come up with a number of mitigation measures. Adversarial training is seen as a way of making models more resistant to manipulation by introducing them to highly engineered malicious inputs when the model is in a learning state. On top of that defensive architectures, e.g., ensemble learning, preprocessing of the inputs, and robust feature selection can be used to decrease the vulnerability of the model to adversarial perturbations. The other important method is Explainable AI (XAI), which increases the transparency of the model by offering an interpretation of a model decision-making process. XAI does not only assist the cybersecurity community to comprehend and justify AI-generated alerts but also instill trust and responsibility in the implementation of AI knowledge concerning risk management and organizing activities within the organization [7].

F. Ethical, Privacy, and Governance Considerations

Implications of AI implementation in cybersecurity are associated with serious ethical, privacy, and regulatory issues [26]. The AI models necessitate the availability of extensive amounts of information, which in many instances contains sensitive personal, organizational or national data. To guarantee the user confidentiality and avoid unauthorized exploitation, it is critical to ensure that the collection, storage, and processing of data adhere to the privacy legislation, including GDPR, HIPAA, or local regulations. In addition to privacy, responsible AI adoption requires the observance of such principles as transparency, fairness, and accountability. The organizations should have a way to describe AI-informed decisions, especially in high stakes security, to warrant measures like threat reduction, access control, or autonomous incident resolution [3]. Ethical governance frameworks and regulatory compliance plans are thus important to prevent abuse, discrimination or unwanted effects of AI implementation [50]. Furthermore, AI-related research, in cooperation with cybersecurity specialists, and policymakers continue to gain importance to strike a balance between innovation and social protection. This type of collaboration can guarantee that

AI technologies would make the process of cybersecurity more effective and defend the rights of the users, facilitate the achievement of fair decision-making, and retain the trust of the population [46], [49]. Through combining ethical control, legal adherence and technical transparency, organizations will be able to implement AI-supported security mechanisms that are efficient and socially accountable.

III. Methodology

A. Systematic Literature Review

This study follows a systematic literature review (SLR) methodology consisting of three stages: planning, conducting, and reporting. To ensure comprehensive and high-quality coverage of the literature, peer-reviewed articles were retrieved from Scopus, Web of Science (WoS), IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink. These databases were selected due to their strong reputation for indexing high-impact journals and conference proceedings in artificial intelligence and cybersecurity. The review focuses on publications from 2017 to 2025, ensuring the inclusion of recent and relevant advances in AI-driven cybersecurity research.

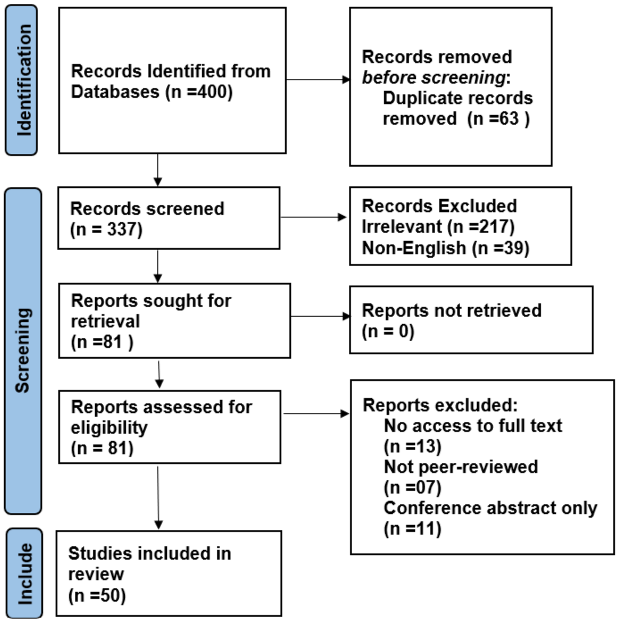
Identification
Screening
Include
Records Identified from Databases (n =400)
Records removed *before screening*: Duplicate records removed (n =63)
Records screened (n = 337)
Records Excluded Irrelevant (n =217) Non-English (n =39)
Reports sought for retrieval (n =81)
Reports not retrieved (n = 0)
Reports assessed for eligibility (n = 81)
Reports excluded: No access to full text (n =13) Not peer-reviewed (n =07) Conference abstract only (n =11)
Studies included in review (n =50)

Fig. 1: PRISMA Diagram

The study selection process followed the PRISMA 2020 guidelines and distinguished between records identified through database searching and those identified through additional methods. Database searches across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink yielded an initial set of records, from which 63 duplicate records were removed, resulting in 337 unique records for screening. During the title and abstract screening stage, 217 records were excluded due to irrelevance and 39 records were excluded for being non-English, leaving 81 records for full-text assessment. Of these, 31 records were excluded (13 inaccessible full texts, 7 non–peer-reviewed publications, and 11 conference abstracts), resulting in 50 studies included from database searches. In addition, snowballing of reference lists and targeted searches of influential authors and research groups were conducted, but all relevant studies identified through these methods were already captured within the final set of 50 included studies. Therefore, no additional unique studies were added beyond those identified through database searching.

B. Research Questions

TABLE I
RESEARCH QUESTIONS

No	Research Question
RQ1	How can artificial intelligence (AI), including machine learning (ML) and deep learning (DL) techniques, enhance the effectiveness and efficiency of cybersecurity systems for critical infrastructure sectors, such as transportation, energy, healthcare, and other mission-critical digital systems?
RQ2	With an emphasis on the transparency, accountability, and reliability requirements for AI-assisted security decisions, what are some of the main obstacles and potential solutions in the development and deployment of explainable AI in cybersecurity applications?

A mapping study is primarily composed of research questions. The research questions that this study attempted to answer are discussed in Table 1. Through these chosen research questions, the mapping study is able to find out research gaps that exist in the current literature.

The major backbone of the mapping study is research questions. The following research questions can be described in Table 1 that this study attempted to answer. Through these few research questions, the mapping study will be capable of finding the gaps in research that exist in the literature.

C. Search Strategy and Reproducibility

To ensure full reproducibility of the systematic literature review, explicit search strings, searched fields, and search dates were documented. Literature searches were conducted on 15 March 2025 across Scopus, Web of Science (WoS), IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink.

The primary search string used was:

("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI") AND (cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")

This search string was adapted to the syntax of each database while preserving semantic equivalence. Searches were applied to the Title, Abstract, and Keywords fields where supported by the database. In IEEE Xplore and ACM Digital Library, searches were performed on Metadata (Title, Abstract, Index Terms).

All retrieved records were exported and documented prior to duplicate removal, screening, and eligibility assessment.

D. Boolean Search Strings and Database-Specific Queries

To ensure transparency and full reproducibility of the systematic literature review, complete Boolean search strings, database-specific query adaptations, and search dates are explicitly reported. Searches were conducted between 12 March 2025 and 15 March 2025 across all selected bibliographic databases. While the logical structure of the search strategy remained consistent, minor syntactic adjustments were applied to accommodate database-specific indexing and query rules.

The core Boolean search string used across databases was:

("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI")
AND
(cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")

TABLE II
DATABASE-SPECIFIC SEARCH CONFIGURATION AND DATES

Database	Exact Query Syntax	Searched Fields	Search Date
Scopus	TITLE-ABS-KEY("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI") AND TITLE-ABS-KEY(cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")	Title, Abstract, Keywords	12 Mar 2025
Web of Science	TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI") AND TS=(cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")	Topic (Title, Abstract, Keywords)	12 Mar 2025
IEEE Xplore	("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI") AND (cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")	Metadata (Title, Abstract, Index Terms)	13 Mar 2025
ACM Digital Library	("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI") AND (cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")	Title, Abstract, Keywords	13 Mar 2025
ScienceDirect	("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI") AND (cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")	Title, Abstract, Keywords	14 Mar 2025
SpringerLink	("artificial intelligence" OR "machine learning" OR "deep learning" OR "explainable AI" OR "XAI") AND (cybersecurity OR "intrusion detection" OR malware OR "threat intelligence" OR "IoT security")	Title, Abstract, Keywords	15 Mar 2025

E. Study Quality Appraisal and Risk of Bias Assessment

To enhance the reliability and validity of the systematic literature review, a study quality appraisal was conducted for all included publications. Given the heterogeneous nature of AI-driven cybersecurity research—encompassing surveys, experimental studies, frameworks, and applied systems—a qualitative assessment approach was adopted rather than a statistical risk-of-bias model. Each study was evaluated based on the following quality criteria:

(1) clarity of research objectives and problem definition;
(2) appropriateness of the AI techniques employed;
(3) transparency of methodology and experimental setup;
(4) relevance of datasets or evaluation scenarios; and
(5) validity of results and conclusions.

Studies that lacked methodological clarity, empirical justification, or peer-review status were excluded during the selection process. This quality appraisal ensured that only methodologically sound and relevant studies contributed to the final synthesis, while minimizing potential bias arising from weak experimental design, incomplete reporting, or unsupported claims.

F. Study Selection

1) Terms and search strings

In this research, the main search will be done via search strings in scientific databases and manual searching by using the relevant keywords. The search strategies are structured based on the population, intervention, comparison, and outcome framework to develop effective search strategies. This structure is naturally determined by the research topics whereby each part provides keywords to the search strings. These search strings were adapted to the syntax and indexing rules of Scopus, Web of Science, IEEE Xplore, and ACM Digital Library to ensure consistent retrieval across databases.

TABLE III
SEARCH TERMS OF THE MAPPING STUDY ON ARTIFICIAL INTELLIGENCE AND SECURITY

Area	Search Terms
Artificial Intelligence in Cybersecurity	"Threat Detection in AI Systems" "AI Incident Response Systems"
Explainable AI (XAI)	"XAI Improves Transparency" "Accountability with XAI Models"
Artificial Intelligence for National Security	"Cyber Warfare through AI Technology" "AI Information Warfare Tools"
AI in IoT Security	"IoT Intrusion Detection using AI" "Security Solutions for IoT Applying AI"

2) Sources

This systematic literature review was conducted using established and reputable bibliographic databases to ensure the quality and credibility of the selected studies. The primary sources included Scopus, Web of Science (WoS), IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink. These databases provide extensive coverage of peer-reviewed journal articles and conference proceedings in the fields of artificial intelligence, machine learning, and cybersecurity, ensuring a robust and comprehensive review of the existing literature.

3) Inclusion and exclusion criteria

This study was designed in terms of selection based on two inclusion criteria and five exclusion criteria. The inclusion and exclusion criteria that were used in the process of filtering are presented in tables 3 and 4 respectively.

The next stage involved a detailed full-text review of the remaining 44 publications to evaluate their relevance and completeness in accordance with the predefined inclusion and exclusion criteria. During this phase, 30 publications were excluded for the following reasons: 7 studies failed to satisfy Inclusion Criteria IC1 and IC2, as they did not adequately address artificial intelligence applications in cybersecurity or discuss relevant theoretical or practical implications; 4 publications were non-primary studies, such as editorials or keynote summaries (EC1); and 19 studies were excluded under Exclusion Criterion EC4, as they primarily provided historical overviews of artificial intelligence or cybersecurity without addressing current or practical applications. Consequently, 14 studies were retained for further analysis.

In the next stage, a backward snowballing approach was employed by examining the reference lists of the 14 retained publications, which resulted in the identification of 11 additional studies. Following rigorous screening based on IC1–IC2 and EC1–EC5, 9 studies were deemed relevant and included.

The next stage focused on identifying influential research groups and leading contributors in the field of AI-driven cybersecurity. A targeted analysis of publications from five prominent researchers or research groups led to the inclusion of four additional high-quality studies, all of which met the established inclusion criteria and did not violate any exclusion conditions. Overall, this systematic multi-stage selection process resulted in a final set of 50 publications, which were included in the comprehensive synthesis and analysis.

TABLE IV
INCLUSION CRITERIA OF THE SELECTION PROCESS

No	Inclusion criteria (IC)
IC1	Studies that investigate artificial intelligence applications for cybersecurity in critical infrastructure or mission-critical environments, including but not limited to transportation, energy, healthcare, IoT ecosystems, enterprise systems, and national security contexts.
IC2	They should discuss theoretical or practical issues, as well as potential opportunities or ramifications for AI-integrated cybersecurity solutions.

TABLE V
EXCLUSION CRITERIA OF THE SELECTION PROCESS

No	Exclusion criteria (EC)
EC1	Documents that only address the legal or regulatory frameworks pertaining to artificial intelligence, without mentioning specific cybersecurity applications, will be excluded.
EC2	Studies that primarily engage in philosophical or ethical discussions regarding AI sentience or consciousness will not be considered.
EC3	Research that mainly examines hardware architectures or detailed technical implementations of AI systems, without relevance to cybersecurity applications, will be excluded.
EC4	Articles that provide only a historical review of the evolution of cybersecurity or artificial intelligence, without discussing current or practical applications, will not be included.
EC5	Studies focusing on artificial intelligence applications beyond the scope of cybersecurity, such as image processing or natural language processing, will be excluded.

TABLE VI
FEATURE EXTRACTION AND SYSTEM RESULTS ACROSS STUDIES

Research Focus Group	Reference No(s).	Feature Extraction	System / Technique Used	Key Results / Outcomes
IoT Security & Smart Ecosystems	[1], [2], [8], [18], [23], [36]	IoT traffic patterns, device behavior, sensor and ecosystem metrics	AI-based anomaly detection, IoT countermeasures, ecosystem validation	High detection accuracy, reduced vulnerabilities, resilient IoT environments
Intrusion Detection & Network Security	[9], [10], [17], [21], [30], [41], [43], [45], [49]	Network packets, protocol features, logs, flow statistics	AI-based IDS, deep learning, real-time threat intelligence	Improved intrusion detection, low false positives, rapid response
Deep Learning, NLP & Generative AI in Cybersecurity	[5], [6], [12], [14], [37], [46]	Deep features, NLP-derived log features, generative AI representations	CNN, RNN, ANN, NLP-based AI, Generative AI models	High classification accuracy, predictive threat detection, adaptive cyber defense
Explainable, Trustworthy & Adversarial AI	[3], [7], [22], [32], [42]	Explainability indicators, compliance metrics, adversarial patterns	XAI frameworks, trustworthy AI models, robustness evaluation	Transparent decisions, improved robustness, secure AI deployment
Enterprise, SOC & Organizational Cybersecurity	[4], [13], [16], [24], [28], [35], [44], [47], [48]	Enterprise logs, SOC metrics, behavioral and system indicators	AI-driven SOC analytics, security management systems	Improved organizational security posture and operational efficiency
National, Military & Psychological Security	[11], [15], [19], [26], [27], [38], [50]	National security events, military network data, psychological behavior features	AI-based monitoring and defense intelligence	Enhanced situational awareness and citizen security
Emerging & Specialized AI Security Applications	[20], [25], [29], [31], [33], [34], [39], [40]	Vision data, chatbot interactions, cybercrime and blockchain patterns	Vision-based AI, AI chatbots, blockchain–AI integration	Enhanced monitoring, cybercrime reduction, secure digital systems

The following table summarizes 50 research studies that investigate the application of Artificial Intelligence (AI) in cybersecurity across diverse domains. The reviewed works cover a broad range of security areas, including intrusion detection systems, IoT and smart ecosystem security, enterprise and SOC operations, malware and threat intelligence, national and military security, explainable and trustworthy AI, as well as emerging applications such as generative AI and blockchain-integrated security systems. The table systematically presents the feature extraction methods, AI techniques employed, and key system-level outcomes, highlighting how AI-driven approaches enhance detection accuracy, reduce false positives, improve real-time threat response, and strengthen overall security resilience. Dominant trends across the literature include the extensive use of machine learning, deep learning models (e.g., CNNs, RNNs, ANNs), natural language processing, and explainable AI (XAI) for anomaly detection, threat classification, situational awareness, and decision support. Furthermore, several studies emphasize transparency, robustness, and ethical considerations to improve trust in AI-based cybersecurity systems.

TABLE VII
SECURITY-RELATED PLATFORMS AND KEY FEATURE

Security Services	Key features
Intrusion Detection Systems (IDS)	Artificial intelligence is employed to detect malicious activities by identifying anomalies that deviate from normal network behavior.
Threat Intelligence Platforms	These platforms gather and analyze threat-related data from multiple sources to produce actionable insights into emerging attack methods and threat trends.
AI-Enhanced Security Control Software	Convolutional Neural Networks (CNNs) are used in facial recognition systems to support access control mechanisms and identity verification.
Cybersecurity Chatbots	Natural Language Processing (NLP) enables chatbots to interpret user inquiries effectively and deliver relevant guidance on information security.

The TABLE VI Anomaly detection within Intrusion Detection Systems in order to recognize unusual network behavior. Effective means of threat prediction through data analytics that allow preventive security measures through Threat Intelligence Platforms. Enhanced access control and identity verification using facial recognition have been included in AI-Enhanced Security Control Software. Technologies such as natural language processing make automated responses possible much like user interaction in Cybersecurity Chatbots.

IV.RESULTS AND DISCUSSION

The research review of 50 articles on AI role in cybersecurity has helped to very clearly realize that it is an exceedingly salient and coherent pattern: AI is rapidly changing the nature of security practices and has provided a great deal of potential to enhance the robustness of Internet cyber defenses. The fact that it can handle enormous uptakes of data, identify complex patterns and react to dynamic threats is still invaluable and as probably most evident in the context of intrusion detection systems (IDS), where AI surpasses the ability of traditional, rule-based ones [4],[8],[11],[18].

The study confirms the effectiveness of the different machine learning algorithms including the Support Vector Machines (SVMs) [5],[8] Random Forest [8],[21] and Artificial Neural Networks (ANNs) [18],[21]. Apparently these methods are effective in detecting indeed malicious activities on networks. Furthermore, the use of deep learning methods, particularly Convolutional Neural Networks (CNNs) [4],[8],[10],[22] are not going to be implemented without difficulties, notwithstanding all the mentioned benefits and the potentially beneficial alterations that AI will introduce to the domain of cybersecurity. Firstly, most AIs are opaque or black-boxed networks beneath a given layer [7]. This leads to suspicion about the determination and responsibility in the ability of security analysts to clarify the actuality of AI-based decisions [7],[23]. This also explains the need to make Explainable AI (XAI) a research priority. XAI is the act of rendering AI systems transparent and explainable by adding the decision-making process of such systems [7].

Cultural There are many XAI techniques under development and research. Such model-agnostic techniques like LIME [7] and SHAP [7] are examples of techniques that can be applied to any type of AI model. A few of such model-specific approaches are the ones that are specific to specific AI architectures. Other uses of XAI not involving intrusion detection [7] are malware analysis [1],[2],[7] phishing detection, and countering social media manipulation [11],[23],[24]. The rise of the XAI is a change of human machine collaboration in cybersecurity. It similarly enhances the clarity and explainability of an artificial intelligence (AI) to enable security professionals to enjoy more benefits of AI without sacrificing a human factor and decision-making on critical security decisions.

V. CONCLUSION

The cyberspace security revolution promised by the contemporary Artificial Intelligence (AI) has been radical and extensive. The recent research and applications have signaled a massive change in the perception of cybersecurity practices, which focuses on the conventional reactive response to the threats, as an active, smart system, capable of forecasting and responding to threats. The automation of complex processes includes intrusion detection, prediction of threats, classification of malware, and automated response to incidents which have been possible with AI technologies. Successive machine learning algorithms, such as Support Vector Machines, random forests, or artificial neural networks, have repeatedly shown high levels of performance in choosing anomalous patterns and malicious actions than modern signature-based or rule-based systems. These algorithms are able to handle flow of big amounts of network data and detect small patterns that may be attacks and give early early warnings, which decreases the time between discovery and response. Machine learning, in particular, deep learning has reinforced the cybersecurity defenses even more. Convolutional Neural Networks, Recurrent Neural Networks, and Generative Adversarial Networks are among the techniques that have been specifically useful in processing complex and high-dimensional data that help systems detect advanced attacks that are usually difficult to detect using the conventional methods of detection. These models learn historical attack patterns by analyzing the logs of large datasets of past cyberattacks, thereby recognizing patterns that have never been observed before, predicting vectors of attacks, and can even aid automated recovery plans. They are critical especially in large scale environments that may have Internet of Things networks, cloud infrastructures, and essential national security systems. In spite of these developments there are major challenges with the adoption of AI in cybersecurity. Among them, one of the limitations is the black box or opaque nature of most AI and deep learning models, in which decision making processes cannot be easily interpreted. Explainable AI methods are also being developed in an attempt to overcome this problem, through offering human-comprehensible clarifications concerning the choices made by AI, and consequently enhance trust, responsibility, and successful human-machine cooperation. Another issue is adversarial attacks, which malicious actors knowingly use to mislead AI models with the intention of manipulating input data. These threats need to be addressed by strong adversarial training, life-cycle testing of models and creating resilient architectures that may resist advanced attacks. The resource needs are also not trivial since training and deployment of sophisticated AI models can be costly and energy-intensive and therefore may not be accessible to smaller organizations. Another important parameter in AI-based cybersecurity is datasets. Good-quality, labeled datasets have played a significant role to both train and validate AI models. Nevertheless, the quick development of cyber threats requires constant changes and additions to such datasets, and taking into account ethical and privacy issues in the collection and use of data. The use of AI should be deployed with transparency, accountability, and regulatory compliance mostly on sensitive issues of responsibility in matters of national security, healthcare, and critical infrastructures. To sum up, AI has the revolutionary potential in cybersecurity, and it is much more powerful than conventional defenses. Combined efforts of machine learning and deep learning allow to detect the threats in advance, perform automatic reaction, and analyze the intricate data faster. Nevertheless, the difficulties are in the form of explain ability, adversarial vulnerability, computational requirements, and ethical issues. The future study should be aimed at the creation of trustworthy and interpretable AI models, the growth and curation of cybersecurity data, and the cooperation of the AI community with cybersecurity specialists and policymakers. With these challenges overcome, AI has the potential to be a pillar in the establishment of a safe, resilient, and adaptable cyber ecosystem, which is capable of resisting the emerging cyber-attacks on organizations, individual, and even individual countries.

References

[1] M. Kuzlu, C. Fair, and O. Guler, “Role of Artificial Intelligence in the Internet of Things (IoT) cybersecurity,” Discover Internet of Things, vol. 1, no. 1, p. 7, 2021, doi: 10.1007/s43926-020-00001-4.

[2] C. Iwendi, S. U. Rehman, A. R. Javed, S. Khan, and G. Srivastava, “Sustainable security for the Internet of Things using artificial intelligence architectures,” ACM Transactions on Internet Technology, vol. 21, no. 3, pp. 1–22, 2021, doi: 10.1145/3448614.

[3] T. Katulić, “Towards the trustworthy AI: Insights from the regulations on data protection and information security,” Medijska Istraživanja, vol. 26, no. 2, pp. 9–28, 2021, doi: 10.22572/mi.26.2.1.

[4] S. Sadik, M. Ahmed, L. F. Sikos, and A. K. M. N. Islam, “Toward a sustainable cybersecurity ecosystem,” Computers, vol. 9, no. 3, p. 74, 2020, doi: 10.3390/computers9030074.

[5] P.Karthika, P.Vidhya Saraswathi, “Image Security Performance Analysis for SVM and ANN Classification Techniques,” International Journal of Recent Technology and Engineering, vol. 8, no. 4S2, pp. 436–442, 2019, doi: 10.35940/ijrte.D1096.1284S219.

[6] X. Liu, Z. Li, Z. Tang, X. Zhang, and H. Wang, “Application of artificial intelligence technology in electromechanical information security situation awareness system,” Scalable Computing: Practice and Experience, vol. 25, no. 1, pp. 127–136, 2024, doi: 10.12694/scpe.v25i1.2280.

[7] Z. Zhang, H. A. Hamadi, E. Damiani, C. Y. Yeun, and F. Taher, “Explainable artificial intelligence applications in cyber security: State-of-the-art in research,” IEEE Access, vol. 10, pp. 93104–93139, 2022, doi: 10.1109/ACCESS.2022.3204051.

[8] S. Zaman et al., “Security threats and artificial intelligence-based countermeasures for Internet of Things networks: A comprehensive survey,” IEEE Access, vol. 9, pp. 94668–94690, 2021, doi: 10.1109/ACCESS.2021.3089681.

[9] S. Gupta, A. S. Sabitha, R. Punhani, et al., “Cyber security threat intelligence using data mining techniques and artificial intelligence,” IJRTE, vol. 8, no. 3, pp. 6133–6140, 2019, doi: 10.35940/ijrte.C5675.098319.

[10] B. Sharma and D. R. Mangrulkar, “Deep learning applications in cyber security: A comprehensive review, challenges and prospects,” International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 08, pp. 148–159, 2019, doi: 10.33564/IJEAST.2019.v04i08.023.

[11] M. N. Al-Suqri and M. Gillani, “A comparative analysis of information and artificial intelligence toward national security,” IEEE Access, vol. 10, pp. 64420–64434, 2022, doi: 10.1109/ACCESS.2022.3183642.

[12] A. Ali, “AI-powered shield: Transforming cybersecurity with neural network innovations,” JMIR Preprints, Nov. 2024, doi: 10.2196/preprints.69000.

[13] S. Thapaliya and A. Bokani, “Leveraging artificial intelligence for enhanced cybersecurity: insights and innovations,” SADGAMAYA, vol. 1, no. 1, pp. 46–52, 2024, doi: 10.3126/sadgamaya.v1i1.66888.

[14] W. S. Ismail, “Threat detection and response using AI and NLP in cybersecurity,” Journal of Internet Services and Information Security, vol. 14, no. 1, pp. 195–205, 2024, doi: 10.58346/JISIS.2024.I1.013.

[15] A. J. Wiranata, “Utilization of artificial intelligence in monitoring and mitigating national security threats,” Archives of Current Research International, vol. 24, no. 11, pp. 23–47, 2024, doi: 10.9734/acri/2024/v24i11945.

[16] K. Lisovskyi and G.-D. Rochenovich, “Artificial intelligence in the security system of enterprise,” Grail Science, no. 27, pp. 308–316, 2023, doi: 10.36074/grail-of-science.12.05.2023.049.

[17] C. Bhatt, “Improving intrusion detection systems with artificial intelligence: A review of techniques and applications,” TURCOMAT, vol. 10, no. 2, pp. 1068–1074, 2019, doi: 10.17762/turcomat.v10i2.13627.

[18] K. Naithani, “AI-based intrusion detection system for Internet of Things (IoT) networks,” TURCOMAT, vol. 10, no. 2, pp. 1095–1100, 2019, doi: 10.17762/turcomat.v10i2.13631.

[19] M. Bistron and Z. Piotrowski, “Artificial intelligence applications in military systems and their influence on sense of security of citizens,” Electronics, vol. 10, no. 7, p. 871, 2021, doi: 10.3390/electronics10070871.

[20] “Vibration guided automatic vision for enhanced security,” ,” Int. J. Innov. Technol. Explor, Eng, vol. 9, no. 2S, pp. 301–306, 2019, doi: 10.35940/ijitee.B1076.1292S19.

[21] S. Hidalgo-Espinoza, K. Chamorro-Cupuerán, and O. Chang-Tortolero, “Intrusion detection in computer systems by using artificial neural networks with deep learning approaches,” CS & IT, pp. 1–12, 2020, doi: 10.5121/csit.2020.101501.

[22] I. Y. Tyukin, D. J. Higham, and A. N. Gorban, “On adversarial examples and stealth attacks in artificial intelligence systems,” arXiv, 2020, doi: 10.48550/ARXIV.2004.04479.

[23] G. Hatzivasilis, N. Papadakis, I. Hatzakis, S. Ioannidis, and G. Vardakis, “Artificial intelligence-driven composition and security validation of an Internet of Things ecosystem,” Applied Sciences, vol. 10, no. 14, p. 4862, 2020, doi: 10.3390/app10144862.

[24] Y. Wu, “Integration of information security and artificial intelligence implementation,” IOP Conf. Ser.: Earth Environ. Sci., vol. 428, no. 1, 012098, 2020, doi: 10.1088/1755-1315/428/1/012098.

[25] S. Hamad and T. Yeferny, “A chatbot for information security,” arXiv, 2020, doi: 10.48550/ARXIV.2012.00826.

[26]T. M. Kolade, “Artificial intelligence and global security: Strengthening international cooperation and diplomatic relations,” Archives of Current Research International, vol. 24, no. 6, pp. 45–56, 2024, doi: 10.9734/acri/2024/v24i61043

[27] A. Averkin, D. Bazarkina, K. Pantserev, and E. Pashentsev, “Artificial intelligence in the context of psychological security: theoretical and practical implications,” in Proc. EUSFLAT, 2019, doi: 10.2991/eusflat19.2019.16.

[28] J. Jain, “Artificial intelligence in the cyber security environment,” in Artificial Intelligence and Data Mining Approaches in Security Frameworks. Wiley, 2021, doi: 10.1002/9781119760429.ch6.

[29] M. Chandra, “Reduction of cyber crimes by effective use of artificial intelligence techniques,” Int. J. Recent Technol. Eng, vol. 8, no. 4, pp. 8643–8645, 2019, doi: 10.35940/ijrte.D8566.118419.

[30] S.-F. Wen, A. Shukla, and B. Katt, “Artificial intelligence for system security assurance: A systematic literature review,” , Int. J. Inf. Secur vol. 24, no. 1, 2025, doi: 10.1007/s10207-024-00959-0.

[31] “Cyber security affairs in empowering technologies,” Int. J. Innov. Technol. Explor. Eng, vol. 8, no. 10S, pp. 1–7, 2019, doi: 10.35940/ijitee.J1001.08810S19.

[32] Y. Wang and T. Nakachi, “Towards secured and transparent artificial intelligence technologies in hierarchical computing networks,” NTT Technical Review, vol. 17, no. 9, pp. 19–26, 2019, doi: 10.53829/ntr201909ra1.

[33] K. Srivastava, “A new approach of artificial intelligence (AI) to cyber security,” International Journal of Research Advent Technology, vol. 7, no. 1, pp. 410–412, 2019, doi: 10.32622/ijrat.71201980.

[34] M. F. Ansari, B. Dash, P. Sharma, and N. Yathiraju, “The impact and limitations of artificial intelligence in cybersecurity: A literature review,” Int. J. Adv. Res. Comput. Commun. Eng, vol. 11, no. 9, 2022, doi: 10.17148/IJARCCE.2022.11912.

[35] R. Raimundo and A. Rosário, “The impact of artificial intelligence on data system security: A literature review,” Sensors, vol. 21, no. 21, p. 7029, 2021, doi: 10.3390/s21217029.

[36] Y. Zhou and Y. Liang, “Application of artificial intelligence technology in network security,” Highlights in Science Engineering and Technology, vol. 92, pp. 479–485, 2024, doi: 10.54097/1mrvaw84.

[37] S. A. M. Ramadhan, S. N. A. Jabbar, and O. A. M. Ramadhan, “Application of artificial intelligence to employ mathematical relations between numbers and their ranks in information security,” Technium, vol. 12, pp. 45–52, 2023, doi: 10.47577/technium.v12i.9360.

[38] T. T. Nguyen, K. D. Nguyen, and X. T. Duong, “Application of artificial intelligence to build a security control software system in local military units,” HIU Journal of Science, vol. 4, pp. 117–124, 2023, doi: 10.59294/HIUJS.VOL.4.2023.394.

[39] A. Jumagaliyeva et al., “Identifying patterns and mechanisms of AI integration in blockchain for e-voting network security,” Eastern-European Journal of Enterprise Technologies, vol. 4, no. 130, pp. 6–18, 2024, doi: 10.15587/1729-4061.2024.305696.

[40] M. Gao, “Obstacles and impacts of artificial intelligence in digital security,” Advances in Economics, Management and Political Science, vol. 93, no. 1, pp. 12–18, 2024, doi: 10.54254/2754-1169/93/20241092.

[41] K. Ovabor, I. O. Sule-Odu, T. Atkison, A. T. Fabusoro, and J. O. Benedict, “AI-driven threat intelligence for real-time cybersecurity: Frameworks, tools, and future directions,” Open Access Research Journal of Science & Technology, vol. 12, no. 2, pp. 040–048, 2024, doi: 10.53022/oarjst.2024.12.2.0135.

[42] O.-M. C. Osazuwa and M. O. Musa, “The expanding attack surface: Securing AI and machine learning systems in security operations,” Int. J. Innov. Sci. Res. Technol, pp. 2498–2505, 2024, doi: 10.38124/ijisrt/IJISRT24MAY1613.

[43] Z. Wang, “Artificial intelligence in cybersecurity threat detection,” Int. J. Comput. Sci. Inf. Technol, vol. 4, no. 1, pp. 203–209, 2024, doi: 10.62051/ijcsit.v4n1.24.

[44] G. Munjal, B. Paul, and M. Kumar, “Application of artificial intelligence in cybersecurity,” in Advances in Information Security, Privacy, and Ethics, IGI Global, 2024, pp. 127–146, doi: 10.4018/979-83693-1431-9.ch006.

[45] S. M. Nour and S. A. Said, “Harnessing the power of AI for effective cybersecurity defense,” in Proc. 6th Int. Conf. Comput. Informatics (ICCI), 2024, pp. 98–102, doi: 10.1109/ICCI61671.2024.10485059.

[46] K. Palani, J. Kethar, S. Prasad, and V. Torremocha, “Impact of AI and generative AI in transforming cybersecurity,” Journal of Studies & Research, vol. 13, no. 2, 2024, doi: 10.47611/jsrhs.v13i2.6710.

[47] S. N. Hasan et al., “The influence of artificial intelligence on data system security,” Int. J. Comput. Eng. Sci. Eng, vol. 11, no. 3, pp. 5589–5598, 2025, doi: 10.22399/ijcesen.3476.

[48] A. Ilyenko, S. Ilyenko, O. Yakovenko, Y. Halych, and V. Pavlenko, “Prospects of integration of artificial intelligence in cybersecurity systems,” Cybersecurity Education Science & Tech., vol. 1, no. 25, pp. 318–329, 2024, doi: 10.28925/2663-4023.2024.25.318329.

[49] G. Poornima and P. R., “Artificial intelligence for cyber security,” in Futuristic Trends in Artificial Intelligence, Vol. 2, Selfypage Developers, 2023, doi: 10.58532/V2BS17P1CH7.

[50] J. Hälterlein, “The use of AI in domestic security practices,” in Handbook of Critical Studies of Artificial Intelligence, Edward Elgar, 2023, pp. 763–772, doi: 10.4337/9781803928562.00077.