International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Emerging Artificial Intelligence Models in Cybersecurity: A Systematic Analysis of Applications, Challenges and Future Directions

Anil Kumar Srivastava[1*], Atul Verma[2], Rajeev Kumar[1, 3]

[1]Center for Multidisciplinary Research and Advanced Studies, Shri Ramswaroop Memorial University, Lucknow-Deva Road, Barabanki-225003, UP, India
anilksrivastava1[at]gmail.com

[2]Department of Computer Science and Information Systems, Shri Ramswaroop Memorial University, Lucknow-Deva Road, Barabanki-225003, UP, India
atulverma.dcsis[at]srmu.ac.in

[3]Department of Computer Science and Engineering, Shri Ramswaroop Memorial University, Lucknow-Deva Road, Barabanki-225003, UP, India
rajeevkr.cse[at]gmail.com

Abstract: ***The increasing sophistication of cyber threats has made securing digital infrastructures, cloud environments, critical systems, and Internet of Things (IoT) networks more challenging. Traditional cybersecurity approaches often struggle to detect advanced attacks and adapt to evolving threat landscapes. Consequently, Artificial Intelligence (AI) has emerged as a powerful enabler of intelligent threat detection, prediction, analysis, and automated response. This study presents a systematic analysis of emerging AI models in cybersecurity, including machine learning, deep learning, transformer-based architectures, Large Language Models (LLMs), generative AI, reinforcement learning, federated learning, explainable AI, and hybrid neuro-symbolic approaches. Using a PRISMA-based methodology, relevant studies were analyzed to evaluate their applications, strengths, limitations, and performance across domains such as intrusion detection, malware analysis, phishing detection, threat intelligence, vulnerability assessment, cloud and IoT security, digital forensics, incident response, and critical infrastructure protection. The analysis further examines AI-driven decision-making techniques, including Multi-Criteria Decision-Making (MCDM) methods and fuzzy logic for cybersecurity risk assessment and threat prioritization. A comparative analysis is conducted based on performance, explainability, scalability, computational efficiency, and trustworthiness. Key challenges, including adversarial attacks, data quality issues, model transparency, privacy concerns, regulatory compliance, and human–AI collaboration, are also discussed. Based on identified research gaps, an AI–Fuzzy Cybersecurity Decision-Making Framework is proposed to enhance cybersecurity governance and operational decision support. Finally, the study highlights future directions such as explainable AI, generative cyber defense, autonomous security operations centers, privacy-preserving federated learning, quantum-AI security, and human-centered security platforms, providing valuable insights for researchers, practitioners, and policymakers.***

Keywords: Artificial Intelligence, Cybersecurity, Large Language Models, Fuzzy Logic, Multi-Criteria Decision-Making.

1. Introduction

The increasing dependence of modern societies on digital technologies has transformed cybersecurity into a critical global concern. Governments, businesses, healthcare institutions, financial organizations, and critical infrastructure providers rely heavily on interconnected systems, cloud computing platforms, IoT devices, and intelligent digital services. While these technological advancements have improved operational efficiency and connectivity, they have also expanded the attack surface available to cybercriminals. Contemporary cyber threats, including ransomware, advanced persistent threats (APTs), phishing campaigns, zero-day exploits, insider threats, and AI-powered cyberattacks, have become increasingly sophisticated, adaptive, and difficult to detect using conventional security mechanisms [1], [2], [3], [4], [5], [6], [7], [8] [9].

Traditional cybersecurity solutions primarily depend on signature-based detection, predefined rules, and manual analysis. Although effective against known threats, these approaches often struggle to identify novel attack patterns, rapidly evolving malware variants, and large-scale cyber incidents occurring in dynamic environments [10]. The growing volume, velocity, and variety of cybersecurity data have further challenged human analysts and security operations centers (SOCs), creating an urgent need for intelligent, scalable, and automated defense mechanisms.

AI has emerged as a transformative technology capable of enhancing cybersecurity capabilities through automated threat detection, predictive analytics, anomaly identification, risk assessment, and intelligent decision-making. Recent advancements in machine learning, deep learning, transformer architectures, LLMs, generative AI, reinforcement learning, federated learning, explainable AI, and neuro-symbolic AI have significantly expanded the scope of AI-driven cybersecurity applications [11], [12], [13]. These technologies enable organizations to proactively identify threats, improve incident response, optimize vulnerability management, and strengthen cyber resilience.

Despite the rapid growth of AI-based cybersecurity research, the existing literature remains fragmented across different AI paradigms and application domains. Furthermore, limited

attention has been given to integrating AI-driven analytics with decision-making methodologies and uncertainty management techniques such as fuzzy logic and MCDM. Consequently, there is a need for a comprehensive analysis that systematically examines emerging AI models, their cybersecurity applications, associated challenges, and future research opportunities while proposing a unified framework for intelligent cybersecurity decision support.

1.1 Motivation and Research Gap

The motivation for this study arises from the increasing complexity of cyber threats and the growing adoption of advanced AI technologies for cyber defense. Recent years have witnessed substantial developments in transformer-based models, generative AI systems, federated learning architectures, explainable AI techniques, and autonomous decision-making frameworks. These innovations have demonstrated promising capabilities for enhancing cybersecurity performance across multiple domains. However, several research gaps remain evident in the existing body of knowledge:

- Most review studies focus on specific AI techniques such as machine learning or deep learning, while overlooking emerging models such as LLMs, generative AI, reinforcement learning, and neuro-symbolic AI.
- Existing surveys rarely provide a unified taxonomy encompassing the full spectrum of emerging AI technologies used in cybersecurity.
- Limited studies investigate the integration of AI models with decision-making techniques.
- The role of fuzzy logic in addressing uncertainty, ambiguity, and risk prioritization within AI-driven cybersecurity systems remains insufficiently explored.
- Comparative evaluations of AI models often emphasize detection accuracy while neglecting explainability, trustworthiness, privacy preservation, computational cost, and scalability.
- Challenges associated with adversarial AI, ethical concerns, regulatory compliance, transparency, and human–AI collaboration are not comprehensively addressed within a single review framework.

1.2 Research Questions

To address the identified research gaps, this study is guided by the following research questions:

- RQ1: What are the major emerging AI models currently being applied in cybersecurity?
- RQ2: How are these AI models utilized across different cybersecurity application domains, including intrusion detection, malware analysis, phishing detection, threat intelligence, vulnerability assessment, and critical infrastructure protection?
- RQ3: What decision-making techniques are integrated with AI systems to enhance cybersecurity risk assessment and strategic decision support?
- RQ4: How does fuzzy logic contribute to cybersecurity decision-making under uncertainty and incomplete information?
- RQ5: What are the comparative strengths and limitations of existing AI-based cybersecurity solutions with respect to accuracy, explainability, scalability, trustworthiness, and computational efficiency?
- RQ6: What challenges and open issues hinder the effective adoption of emerging AI models in cybersecurity environments?
- RQ7: What future research directions and integrated frameworks can support the development of intelligent, adaptive, and trustworthy cybersecurity systems?

1.3 Objectives and Contributions

The primary objective of this study is to systematically review emerging AI models in cybersecurity and provide a comprehensive understanding of their applications, challenges, and future directions. The specific objectives are as follows:

- To identify and classify emerging AI models relevant to cybersecurity applications.
- To analyze the application of AI models across diverse cybersecurity domains.
- To investigate AI-driven decision-making approaches used for cybersecurity risk assessment and management.
- To examine the role of fuzzy logic in cybersecurity decision support systems.
- To compare existing AI-based cybersecurity solutions using multiple performance and trust-related metrics.
- To identify key challenges, limitations, and research opportunities associated with AI-enabled cybersecurity.
- To propose a novel AI–Fuzzy cybersecurity decision-making framework for intelligent threat management and cyber defense.
- The major contributions of this paper include:
- Development of a comprehensive taxonomy of emerging AI models used in cybersecurity.
- Systematic review of AI applications across multiple cybersecurity domains.
- Detailed analysis of AI-driven decision-making and fuzzy logic techniques for cybersecurity governance.
- Comparative evaluation of existing AI-based cybersecurity solutions.
- Identification of critical challenges and open research issues.
- Proposal of an integrated AI–Fuzzy cybersecurity decision-making framework combining threat intelligence, AI analytics, fuzzy inference, and MCDM approaches.
- Presentation of future research directions toward explainable, trustworthy, autonomous, and privacy-preserving cybersecurity systems.

2. Systematic Review Methodology

To ensure rigor, transparency, and reproducibility, this study adopts a systematic literature review (SLR) methodology based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. The methodology was designed to comprehensively identify, evaluate, and synthesize scholarly literature related to emerging AI models and their applications in cybersecurity. The review process involved defining research questions, selecting relevant databases, developing search strategies, applying inclusion and exclusion criteria, conducting a PRISMA-based screening process, and performing qualitative and comparative analyses of the selected studies.

2.1 Review Protocol

A structured review protocol was developed to minimize bias and ensure consistency throughout the review process. The protocol consisted of five major phases:

Phase 1: Research Planning
- Identification of research objectives and research questions.
- Definition of the review scope focusing on emerging AI models in cybersecurity.
- Development of review criteria and evaluation metrics.

Phase 2: Literature Identification
- Selection of relevant academic databases.
- Construction of search strings using cybersecurity and AI-related keywords.
- Retrieval of potentially relevant studies.

Phase 3: Screening and Eligibility Assessment
- Removal of duplicate records.
- Title and abstract screening.
- Full-text eligibility assessment based on predefined criteria.

Phase 4: Data Extraction
- Collection of bibliographic and technical information from selected studies.
- Categorization of studies according to AI model type, cybersecurity application domain, performance metrics, and decision-making approaches.

Phase 5: Data Analysis and Synthesis
- Qualitative synthesis of findings.
- Comparative analysis of AI models.
- Identification of research trends, challenges, and future directions.
- Development of a proposed AI–Fuzzy cybersecurity decision-making framework.

The review protocol was designed to answer the research questions presented in Section 1 while maintaining methodological transparency and reproducibility (Figure 1).

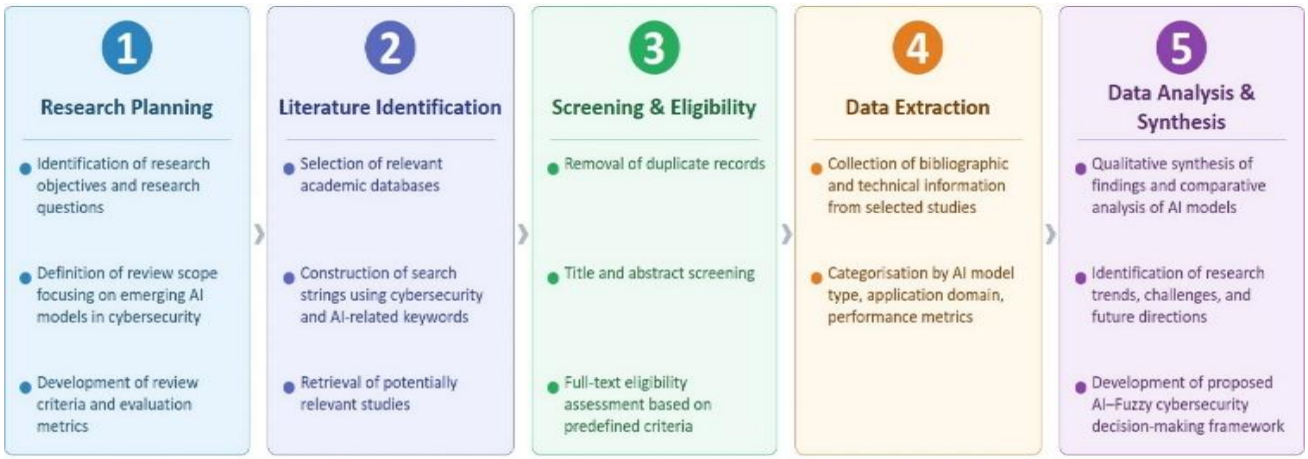

Figure 1: Review Protocol

2.2 Database Selection and Search Strategy

To ensure comprehensive coverage of high-quality research, literature was collected from major scientific databases widely recognized in computer science, cybersecurity, and AI research.

- **Selected Databases:** The literature search was conducted using several leading academic databases, including IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Wiley Online Library, Scopus, Web of Science, MDPI, Taylor & Francis Online, and Google Scholar for supplementary searches. These databases were selected because they provide extensive coverage of high-quality, peer-reviewed journal articles, conference proceedings, review papers, and book chapters relevant to the fields of AI and cybersecurity. The inclusion of multiple databases ensured comprehensive coverage of both foundational and emerging research, thereby enhancing the reliability and completeness of the review process.
- **Search Strategy:** The search process employed a comprehensive set of keywords related to AI, machine learning, cybersecurity, threat intelligence, fuzzy systems, and decision-making techniques. Representative search strings included combinations such as ("Artificial Intelligence" OR "AI") AND ("Cybersecurity" OR "Cyber Security" OR "Information Security"); ("Machine Learning" OR "Deep Learning" OR "Transformer" OR "Large Language Model" OR "Generative AI") AND ("Intrusion Detection" OR "Malware Detection" OR "Threat Intelligence"); ("Federated Learning" OR "Explainable AI" OR "Reinforcement Learning") AND ("Cyber Defense" OR "Risk Assessment"); ("Fuzzy Logic" OR "Neuro-Fuzzy") AND ("Cybersecurity Decision Making" OR "Cyber Risk Management"); and ("AHP" OR "TOPSIS") AND ("Cybersecurity"). These search strings were adapted to the specific syntax requirements of each database to ensure comprehensive retrieval of relevant studies. To capture the latest advancements in AI-driven cybersecurity solutions, the search was restricted to publications published between 2015 and 2026, covering recent developments in emerging AI technologies and their applications in cybersecurity.
- **Search Fields:** The search was conducted across multiple fields, including titles, abstracts, keywords, and

full-text content where supported by the selected databases. Searching across these fields helped maximize the retrieval of relevant studies while minimizing the risk of overlooking significant contributions. To ensure consistency and facilitate a rigorous analysis, only publications written in English were included in the review.

2.3 Inclusion and Exclusion Criteria

To ensure the relevance, quality, and rigor of the selected literature, predefined inclusion and exclusion criteria were established and systematically applied throughout the screening process. Studies were included if they were published between 2015 and 2026, appeared as peer-reviewed journal articles, conference papers, review articles, or book chapters, and focused on the application of AI techniques to cybersecurity challenges. Eligible studies investigated emerging AI approaches such as machine learning, deep learning, transformer-based models, LLMs, generative AI, reinforcement learning, federated learning, explainable AI, or hybrid AI frameworks. Additionally, the studies were required to address cybersecurity domains including intrusion detection, malware detection, phishing detection, threat intelligence, digital forensics, cloud security, IoT security, critical infrastructure protection, cybersecurity decision-making, fuzzy logic, or risk assessment, and to provide performance evaluations, experimental results, or practical implementation insights.

Studies were excluded if they were non-English publications, duplicate records, editorials, short communications, posters, tutorials, or other non-peer-reviewed works. Publications lacking sufficient methodological details, inaccessible full texts, or relevance to cybersecurity applications were also removed. Furthermore, studies focusing solely on traditional statistical methods without incorporating AI techniques, as well as those primarily addressing non-security applications of AI, were excluded. The systematic application of these criteria ensured that only high-quality, relevant, and methodologically sound studies were retained for detailed analysis.

2.4 PRISMA-Based Study Selection

The PRISMA framework was employed to document and report the study selection process systematically.

The screening process consisted of four stages as shown in Figure 2.

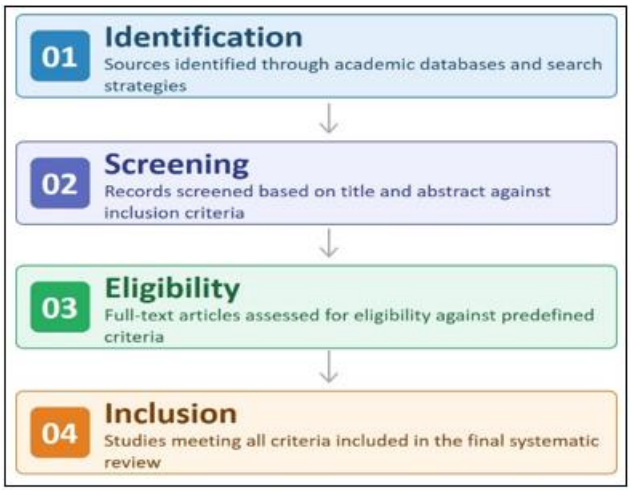

Figure 2: Screening Process

- **Identification:** Literature searches across selected databases initially yielded a large number of records. Additional studies were identified through reference list screening and manual searches.
- **Screening:** Duplicate records were removed using reference management software. Titles and abstracts were then screened against the inclusion and exclusion criteria.
- **Eligibility:** The full text of potentially relevant studies was examined to assess methodological quality, relevance, and completeness.
- **Inclusion:** Studies satisfying all eligibility requirements were included in the final review dataset.

A typical PRISMA workflow for this review may be summarized as Table 1. Further, A PRISMA flow diagram illustrating the selection process is provided in Figure 3.

Table 1: PRISMA Workflow

PRISMA Stage	*Number of Records (Example)*
Records Identified	3,250
Duplicate Records Removed	620
Records Screened	2,630
Records Excluded after Title/Abstract Review	1,850
Full-Text Articles Assessed	780
Full-Text Articles Excluded	480
Studies Included in Final Review	300

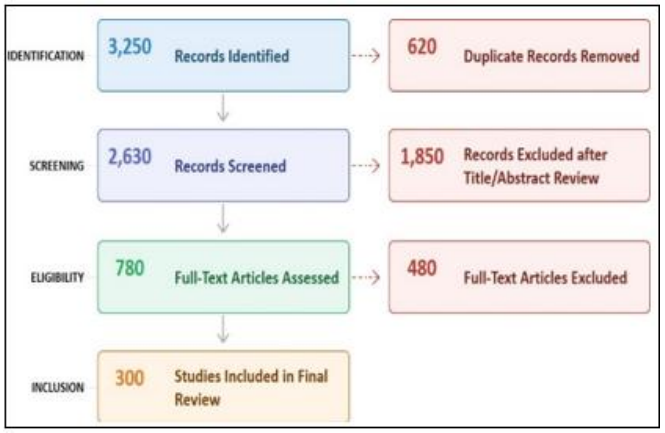

Figure 3: PRISMA Record Flow

2.5 Data Extraction and Analysis

A structured data extraction form was developed to ensure consistency, accuracy, and reproducibility in collecting information from the selected studies. For each study, key attributes were extracted, including authors and publication year, publication source, AI model type, cybersecurity application domain, datasets utilized, evaluation metrics, decision-making techniques employed, fuzzy logic integration, key findings, advantages, limitations, and future research recommendations. To facilitate systematic analysis, the selected studies were classified according to AI categories and cybersecurity domains. The AI categories included Machine Learning, Deep Learning, Transformer-Based Models, LLMs, Generative AI, Reinforcement Learning, Federated Learning, Explainable AI, and Neuro-Symbolic AI. The cybersecurity domains comprised Intrusion Detection, Malware Analysis, Phishing Detection, Threat Intelligence, Vulnerability Assessment, Cloud Security, Edge Security, IoT Security, Digital Forensics, Incident Response, and Critical Infrastructure Security.

The extracted data were analyzed using multiple complementary techniques. Descriptive analysis was conducted to examine publication trends, research distribution across AI categories, and the frequency of cybersecurity application domains. Comparative analysis was performed to evaluate and compare AI-based cybersecurity solutions based on metrics such as accuracy, precision, recall, F1-score, explainability, scalability, and computational complexity. Thematic analysis was employed to identify emerging trends, research challenges, open issues, and future research opportunities within the field. Finally, framework synthesis was carried out by integrating the findings from the reviewed studies to develop a novel AI–Fuzzy Cybersecurity Decision-Making Framework. This methodology provides a comprehensive, systematic, and reproducible approach for evaluating emerging AI models in cybersecurity and supports the development of evidence-based conclusions and future research directions.

3. Taxonomy of Emerging AI Models for Cybersecurity

The increasing sophistication of cyber threats has accelerated the adoption of AI technologies in cybersecurity. Modern cyberattacks are characterized by their dynamic behavior, ability to evade traditional defenses, and exploitation of large-scale digital infrastructures. Consequently, AI has emerged as a critical enabler for intelligent cyber defense by facilitating automated threat detection, anomaly identification, predictive analytics, incident response, and strategic decision-making. Over the past decade, significant advancements have been witnessed in machine learning, deep learning, transformer architectures, generative AI, reinforcement learning, federated learning, explainable AI, and neuro-symbolic intelligence [14], [15]. These developments have expanded the scope of cybersecurity applications while introducing new opportunities and challenges.

To provide a structured understanding of the current landscape, emerging AI models used in cybersecurity can be categorized into eight major classes: Machine Learning Models, Deep Learning Models, Transformer and LLMs, Generative AI Models, Reinforcement Learning Models, Federated Learning Models, Explainable AI Models, and Hybrid and Neuro-Symbolic AI Models. Each category offers distinct capabilities for addressing cybersecurity problems and contributes differently to cyber defense strategies. Figure X illustrates the proposed taxonomy, while Table 2 presents a comprehensive classification of emerging AI models and their cybersecurity applications.

Table 2. Taxonomy of Emerging AI Models for Cybersecurity

AI Category	Representative Models	Core Characteristics	Major Cybersecurity Applications	Advantages	Limitations
Machine Learning (ML)	Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, Logistic Regression, KNN, XGBoost	Learns patterns from labeled or unlabeled data using statistical techniques	Intrusion Detection, Malware Classification, Spam Detection, Phishing Detection, User Behavior Analytics	Interpretable, computationally efficient, effective for structured data	Requires feature engineering, limited adaptability to novel attacks
Deep Learning (DL)	ANN, CNN, RNN, LSTM, GRU, Autoencoders	Learns hierarchical representations from large datasets through multi-layer neural networks	Malware Detection, Network Anomaly Detection, Threat Prediction, Botnet Detection, Traffic Classification	Automatic feature extraction, high detection accuracy	High computational cost, large data requirements, limited explainability
Transformer and LLMs	Transformer, BERT, RoBERTa, GPT, LLaMA, Gemini, Claude, Falcon	Attention-based architectures capable of understanding contextual relationships in data and text	Threat Intelligence, Vulnerability Analysis, Security Report Summarization, Alert Triage, Incident Investigation	Strong contextual reasoning, automation of security workflows	Hallucinations, privacy concerns, resource-intensive training
Generative AI Models	GANs, Variational Autoencoders, Diffusion Models	Generate synthetic data and simulate attack scenarios	Security Dataset Generation, Adversarial Testing, Malware Simulation, Cyber Range Training	Addresses data scarcity, improves robustness evaluation	Potential misuse by attackers, ethical concerns
Reinforcement Learning (RL)	Q-Learning, Deep Q-Networks, Actor-Critic Models, Policy Gradient Methods	Learns optimal actions through interaction with dynamic environments	Adaptive Cyber Defense, Automated Response Systems, Firewall Optimization, Resource Allocation	Autonomous decision-making, continuous adaptation	Complex training, reward engineering challenges

Federated Learning (FL)	Horizontal FL, Vertical FL, Federated Transfer Learning	Enables decentralized learning without sharing raw data	Privacy-Preserving IDS, Collaborative Threat Intelligence, Healthcare and Financial Cybersecurity	Enhanced privacy and regulatory compliance	Communication overhead, vulnerability to poisoning attacks
Explainable AI (XAI)	SHAP, LIME, Integrated Gradients, Attention Visualization	Provides transparent and interpretable explanations for AI decisions	Risk Assessment, Alert Interpretation, Compliance Auditing, Malware Analysis	Improves trust and accountability	Increased complexity and processing overhead
Hybrid and Neuro-Symbolic AI	ML-Fuzzy Systems, DL-RL Models, LLM-Knowledge Graphs, Neuro-Symbolic Architectures	Combines learning, reasoning, and uncertainty handling mechanisms	Cyber Risk Assessment, Threat Intelligence, Incident Response, Security Governance	High explainability, reasoning capability, adaptability	Architectural complexity and scalability challenges

4. Cybersecurity Applications of Emerging AI Models

Emerging AI models have significantly transformed cybersecurity by enabling intelligent, adaptive, and automated defense mechanisms across diverse application domains. As cyber threats become increasingly complex, traditional rule-based systems are often insufficient to detect and mitigate advanced attacks. AI-driven approaches provide enhanced capabilities for real-time analysis, predictive security, and autonomous response. This section presents key cybersecurity application areas where emerging AI models- including machine learning, deep learning, transformers, generative AI, reinforcement learning, federated learning, explainable AI, and hybrid neuro-symbolic systems- are widely deployed.

4.1 Intrusion Detection and Prevention

Intrusion Detection and Prevention Systems (IDPS) represent one of the most mature and widely adopted applications of AI in cybersecurity. Machine learning models such as Random Forest, Support Vector Machines, and XGBoost are commonly used for signature-based and anomaly-based intrusion detection [16]. Deep learning architectures, including CNNs, RNNs, and LSTMs, further enhance detection accuracy by learning complex temporal and spatial patterns in network traffic.

Emerging transformer-based and federated learning approaches enable scalable and privacy-preserving intrusion detection across distributed networks [17]. Reinforcement learning techniques are increasingly used for adaptive intrusion prevention, allowing systems to dynamically respond to evolving threats. Additionally, explainable AI enhances trust by providing interpretable detection results for security analysts.

4.2 Malware Analysis and Detection

Malware analysis has greatly benefited from AI-based techniques capable of identifying both known and unknown malicious software variants. Machine learning models utilize static and dynamic features such as opcode sequences, API calls, and binary structures to classify malware families [18]. Deep learning models, particularly CNNs, are effective in malware image representation analysis, while recurrent networks capture sequential execution patterns.

Generative AI models assist in creating synthetic malware samples for training robust detection systems, whereas transformer-based models support code analysis and malware behavior interpretation [19]. Federated learning enables collaborative malware detection across organizations without sharing sensitive data. Explainable AI techniques improve forensic analysis by clarifying why specific software is classified as malicious.

4.3 Phishing and Social Engineering Detection

Phishing attacks and social engineering remain among the most prevalent cybersecurity threats due to their reliance on human deception. AI-driven natural language processing models, especially transformers and LLMs, are highly effective in detecting phishing emails, malicious URLs, and fraudulent messages [20].

Machine learning classifiers analyze email metadata, domain reputation, and linguistic features, while deep learning models identify hidden semantic patterns in text content. Generative AI is also used to simulate phishing attacks for training and awareness programs. Explainable AI plays a crucial role in helping users and security teams understand phishing detection decisions, thereby improving trust and awareness.

4.4 Threat Intelligence and Threat Hunting

Threat intelligence involves collecting, analyzing, and interpreting data related to cyber threats to support proactive defense strategies. Transformer-based models and LLMs have significantly enhanced threat intelligence by automating the extraction of insights from unstructured data sources such as security reports, blogs, and dark web forums [21].

Machine learning models assist in clustering threat actors and identifying attack patterns, while reinforcement learning supports adaptive threat hunting strategies. Knowledge graph-based and neuro-symbolic AI models integrate structured and unstructured data for advanced reasoning. Federated learning enables collaborative intelligence sharing among organizations while preserving data privacy.

4.5 Vulnerability Assessment and Risk Prediction

AI-based vulnerability assessment systems analyze software systems, networks, and configurations to identify potential weaknesses before exploitation. Machine learning models are widely used for vulnerability scoring and risk

classification, while deep learning approaches analyze code repositories and system logs for hidden vulnerabilities.

Transformer-based models assist in automated code review and CVE (Common Vulnerabilities and Exposures) analysis. Fuzzy logic and MCDM techniques such as AHP, TOPSIS, and VIKOR are increasingly integrated with AI models to support uncertainty-aware risk prioritization [22]. Reinforcement learning further enhances risk prediction by continuously adapting to evolving threat landscapes.

4.6 Cloud, Edge, and IoT Security

The rapid expansion of cloud computing, edge computing, and IoT ecosystems has introduced new cybersecurity challenges related to scalability, heterogeneity, and distributed attack surfaces. AI models play a critical role in securing these environments through real-time monitoring and anomaly detection.

Deep learning models are commonly used for IoT intrusion detection, while federated learning ensures privacy-preserving analytics across distributed devices [23]. Machine learning models support workload classification and anomaly detection in cloud infrastructures. Reinforcement learning enables adaptive resource allocation and dynamic security policy enforcement. Explainable AI ensures transparency in decision-making processes across multi-tenant cloud environments.

4.7 Digital Forensics and Incident Response

Digital forensics involves the identification, preservation, analysis, and presentation of digital evidence. AI techniques enhance forensic investigations by automating log analysis, malware tracing, and anomaly detection. Machine learning models classify forensic artifacts, while deep learning models analyze complex multimedia and system logs.

Transformer-based models assist in summarizing incident reports and correlating multi-source forensic data [24]. Reinforcement learning supports automated incident response strategies, enabling rapid containment and mitigation of cyber incidents. Explainable AI ensures that forensic conclusions are interpretable and legally admissible, improving the reliability of AI-assisted investigations.

4.8 Critical Infrastructure Protection

Critical infrastructure systems, including energy grids, transportation networks, healthcare systems, and financial services, are increasingly targeted by sophisticated cyberattacks. AI-based cybersecurity solutions are essential for ensuring resilience and continuity of these systems.

Machine learning and deep learning models provide real-time monitoring and anomaly detection for industrial control systems (ICS) and supervisory control and data acquisition (SCADA) networks [25]. Federated learning enables secure collaboration across infrastructure operators without exposing sensitive operational data. Reinforcement learning supports adaptive defense strategies for dynamic threat environments. Hybrid and neuro-symbolic AI models integrate domain knowledge with learning capabilities to improve decision-making in high-stakes environments. Explainable AI is particularly critical in this domain, as decision transparency is essential for safety-critical infrastructure operations and regulatory compliance.

5. AI-Driven Decision-Making Techniques in Cybersecurity

Cybersecurity decision-making is inherently complex due to the dynamic, uncertain, and adversarial nature of cyber environments. Security analysts and automated systems must evaluate multiple conflicting criteria such as risk severity, attack likelihood, system vulnerability, cost of mitigation, and operational impact. Traditional rule-based decision systems are often inadequate in such scenarios because they lack the capability to handle uncertainty, interdependencies, and real-time adaptive reasoning [26]. To address these limitations, AI has been increasingly integrated with advanced decision-making methodologies, particularly MCDM, probabilistic reasoning, and game-theoretic models.

5.1 Multi-Criteria Decision-Making Approaches

MCDM methods provide a structured framework for evaluating and selecting optimal alternatives when multiple conflicting criteria are involved. In cybersecurity, MCDM techniques are widely used for risk assessment, vulnerability prioritization, security investment planning, and incident response selection.

AI-enhanced MCDM systems integrate data-driven insights from machine learning models with structured decision frameworks to improve accuracy and robustness [27]. These hybrid approaches allow security analysts to incorporate both quantitative metrics (e.g., attack frequency, detection accuracy) and qualitative factors (e.g., system criticality, expert judgment), thereby enabling more balanced and informed cybersecurity decisions.

5.2 AHP-Based Security Decisions

The Analytic Hierarchy Process (AHP) is one of the most widely used MCDM techniques in cybersecurity decision-making [28]. It decomposes complex decision problems into hierarchical structures consisting of goals, criteria, sub-criteria, and alternatives. Pairwise comparisons are then used to assign relative weights to each factor.

In cybersecurity, AHP is commonly applied to:
- Risk prioritization of vulnerabilities
- Selection of intrusion detection systems
- Evaluation of security controls
- Prioritization of incident response strategies

When integrated with AI models, AHP benefits from data-driven weight calibration, reducing subjectivity and improving consistency in decision-making. However, AHP can become computationally intensive when dealing with large-scale cybersecurity environments involving numerous variables.

5.3 TOPSIS-Based Risk Prioritization

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) and VIKOR are widely used MCDM techniques for ranking and prioritizing cybersecurity risks and solutions.

TOPSIS evaluates alternatives based on their distance from an ideal solution (best case) and a negative ideal solution (worst case), making it suitable for ranking vulnerabilities, threats, and security tools [29]. VIKOR, on the other hand, focuses on identifying compromise solutions that provide a balance between group utility and individual regret.

In cybersecurity applications, these methods are used for:

- Prioritizing vulnerabilities based on severity and exploitability
- Ranking security patches and mitigation strategies
- Selecting optimal cloud security configurations
- Evaluating intrusion detection system performance

AI integration enhances these methods by supplying real-time risk metrics derived from machine learning and deep learning models, thereby improving decision responsiveness in dynamic environments.

AI-driven decision-making techniques significantly enhance cybersecurity by enabling structured, adaptive, and intelligent evaluation of complex risk environments. MCDM methods such as AHP and TOPSIS, provide structured decision frameworks, when integrated with AI models, these techniques collectively enable robust, scalable, and intelligent cybersecurity decision-making systems suitable for dynamic and high-risk digital environments.

6. Fuzzy Logic for Cybersecurity Decision Support

Cybersecurity environments are inherently uncertain due to incomplete information, ambiguous threat indicators, and rapidly evolving attack techniques. Traditional binary decision-making approaches are often inadequate for handling such complexity, making fuzzy logic a valuable tool for cybersecurity applications. By allowing partial membership and linguistic reasoning, fuzzy logic effectively manages uncertainty in risk assessment, threat evaluation, and security decision-making [30]. Fuzzy risk assessment models use qualitative variables such as low, medium, and high to evaluate vulnerabilities, attack likelihoods, and potential impacts, enabling more realistic risk analysis in cloud, IoT, and critical infrastructure environments. Similarly, fuzzy threat prioritization techniques aggregate multiple factors, including severity, exploitability, and business impact, to support accurate ranking of vulnerabilities and cyber threats under uncertain conditions.

Fuzzy logic is also widely integrated with AI, machine learning, and MCDM techniques to enhance cybersecurity intelligence and decision support. Fuzzy-based intrusion detection systems improve anomaly detection by replacing rigid thresholds with flexible membership functions, reducing false positives and improving detection accuracy [31]. Neuro-fuzzy systems further combine neural networks with fuzzy reasoning to provide both learning capability and interpretability for applications such as intrusion detection, malware analysis, and dynamic risk assessment [32]. Additionally, fuzzy-MCDM approaches, including Fuzzy AHP and Fuzzy TOPSIS, support cybersecurity risk prioritization, security solution selection, and strategic decision-making. Together, these approaches provide a robust framework for developing intelligent, adaptive, and uncertainty-aware cybersecurity systems.

7. Comparative Analysis of Existing AI-Based Cybersecurity Solutions

The rapid proliferation of AI techniques in cybersecurity has led to a wide range of proposed solutions addressing intrusion detection, malware analysis, threat intelligence, and risk management. However, these solutions vary significantly in terms of performance, scalability, interpretability, and computational efficiency. A systematic comparative analysis is therefore essential to evaluate their strengths, limitations, and applicability in real-world cybersecurity environments.

This section provides a structured comparison of existing AI-based cybersecurity solutions based on key evaluation metrics, including accuracy, detection capability, explainability, trustworthiness, scalability, and computational cost. The objective is to identify performance trade-offs and highlight research gaps that motivate the development of hybrid AI–fuzzy decision-making frameworks.

7.1 Performance Evaluation Metrics

The performance of AI-based cybersecurity systems is commonly evaluated using a combination of quantitative and qualitative metrics to assess their effectiveness, reliability, and operational suitability. Key metrics include accuracy, precision, recall (detection rate), F1-score, false positive rate (FPR), false negative rate (FNR), and AUC-ROC, which collectively measure the model's ability to correctly identify cyber threats while minimizing misclassifications. In addition to detection performance, operational metrics such as detection latency and computational overhead are also considered to evaluate the speed and resource efficiency of cybersecurity solutions during training and deployment.

A comprehensive evaluation is particularly important in cybersecurity because relying on a single metric, such as accuracy, may provide a misleading assessment of real-world performance. For example, a model with high accuracy may still generate excessive false positives or fail to detect critical attacks. Therefore, cybersecurity researchers and practitioners typically employ a balanced assessment framework that considers detection capability, error rates, response time, scalability, and computational efficiency. Such multidimensional evaluation enables a more realistic comparison of AI models and supports the selection of effective cybersecurity solutions for diverse operational environments.

7.2 Accuracy and Detection Capability

Deep learning and transformer-based models generally demonstrate superior accuracy and detection capability compared to traditional machine learning approaches. Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and ensemble models such as Random Forest and XGBoost are widely reported to achieve high detection rates in intrusion detection and malware classification tasks [33]. Transformer-based architectures and LLMs further enhance detection capability by analyzing unstructured data such as logs, threat reports, and security alerts [34]. However, their performance is highly dependent on training data quality and domain adaptation. Despite their high accuracy, these models may suffer from overfitting, dataset bias, and reduced generalization to unseen attack types. Hybrid models combining deep learning with fuzzy logic or ensemble learning often provide improved robustness in dynamic cybersecurity environments.

7.3 Explainability and Trustworthiness

Explainability is a critical requirement in cybersecurity applications, particularly in high-stakes environments such as critical infrastructure, healthcare systems, and financial networks. Security analysts and decision-makers must understand the rationale behind AI-generated predictions to ensure accountability and compliance. Machine learning models such as Decision Trees and Rule-Based Systems offer high interpretability but limited expressive power. In contrast, deep learning and transformer-based models provide high accuracy but suffer from "black-box" behavior. Explainable AI (XAI) techniques such as SHAP, LIME, and attention visualization have been introduced to bridge this gap by providing post-hoc explanations [35]. However, these methods may not fully capture internal model reasoning and can introduce additional computational complexity. Fuzzy logic-based systems and neuro-fuzzy models offer a natural balance between interpretability and performance by representing decisions in human-readable linguistic rules, making them highly suitable for cybersecurity decision support systems.

7.4 Scalability and Computational Cost

Scalability is a key factor in evaluating cybersecurity solutions, especially in large-scale environments such as cloud computing, IoT networks, and enterprise systems. Traditional machine learning models are generally lightweight and scalable, making them suitable for real-time applications with limited resources [36]. Deep learning models, while highly accurate, require significant computational resources, including GPUs and large memory capacity, which can limit their deployment in edge or resource-constrained environments. Transformer-based and LLMs further increase computational overhead due to their massive parameter sizes. Federated learning provides a scalable and privacy-preserving alternative by distributing computation across multiple nodes; however, it introduces communication overhead and synchronization challenges. Reinforcement learning models may also face scalability issues due to the complexity of training environments. Hybrid architectures that combine lightweight models with fuzzy logic or rule-based systems are increasingly being explored to achieve an optimal balance between scalability and performance.

7.5 Comparative Summary of Reviewed Studies

Table 3 presents a comparative summary of representative AI-based cybersecurity approaches based on key evaluation criteria.

Table 3: Comparative Analysis of AI-Based Cybersecurity Solutions

Approach Category	Example Models	Detection Performance	Explainability	Scalability	Computational Cost	Key Strengths	Key Limitations
Machine Learning	SVM, RF, XGBoost	Moderate to High	High	High	Low	Simple, fast, interpretable	Limited feature learning
Deep Learning	CNN, LSTM, Autoencoders	High	Low to Moderate	Moderate	High	Strong feature extraction	Resource intensive
Transformers / LLMs	BERT, GPT, LLaMA	High	Low	Low to Moderate	Very High	Context-aware analysis	High cost, hallucinations
Generative AI	GANs, VAEs	Moderate	Low	Moderate	High	Data augmentation	Misuse risk
Reinforcement Learning	DQN, Actor-Critic	Adaptive	Low	Moderate	High	Autonomous response	Complex training
Federated Learning	FL frameworks	High (distributed)	Moderate	High	Moderate	Privacy-preserving	Communication overhead
Explainable AI	SHAP, LIME-based models	Moderate	High	Moderate	Moderate	Transparency	Limited real-time use
Hybrid / Neuro-Symbolic	ML + Fuzzy + DL	High	High	Moderate	Moderate to High	Balanced performance	Architectural complexity

The comparative analysis highlights that no single AI approach is sufficient to address all cybersecurity requirements. While deep learning and transformer-based models offer superior detection performance, they often lack explainability and impose high computational costs. In contrast, machine learning and fuzzy logic-based systems provide interpretability but may lack advanced feature representation capabilities. Hybrid approaches, particularly those integrating AI with fuzzy logic and decision-making frameworks, offer a promising direction by balancing

accuracy, interpretability, scalability, and computational efficiency. These findings strongly motivate the development of integrated AI–Fuzzy cybersecurity decision-making frameworks, which are further discussed in the subsequent section.

8. Challenges and Open Issues

Despite significant advancements in AI-driven cybersecurity solutions, several critical challenges and open issues continue to hinder their widespread adoption and reliable deployment in real-world environments. The complexity of modern cyber threats, combined with the limitations of current AI models, introduces concerns related to robustness, trustworthiness, privacy, ethics, and human–AI interaction. This section highlights the key challenges associated with emerging AI models in cybersecurity and identifies open research problems that require further investigation.

8.1 Adversarial AI Attacks

One of the most critical challenges in AI-based cybersecurity systems is their vulnerability to adversarial attacks. Attackers can manipulate input data to deceive AI models into making incorrect predictions, leading to false negatives or false positives in threat detection systems [37]. Techniques such as adversarial malware generation, data poisoning, evasion attacks, and model inversion attacks pose significant risks to machine learning and deep learning-based security systems.
In cybersecurity contexts, adversarial AI can be used to bypass intrusion detection systems, evade malware classifiers, and manipulate threat intelligence systems. Transformer-based models and LLMs are also susceptible to prompt injection, data leakage, and adversarial text manipulation. Ensuring robustness against such attacks remains an open research challenge requiring the development of resilient learning frameworks and adversarial training strategies.

8.2 Data Quality and Availability

High-quality, labeled cybersecurity datasets are essential for training effective AI models. However, obtaining such data is often difficult due to privacy concerns, regulatory restrictions, and the sensitive nature of cyber incidents. Additionally, cybersecurity data is often imbalanced, noisy, incomplete, and highly dynamic, which negatively affects model performance.

Many existing AI models are trained on outdated or synthetic datasets that may not accurately reflect real-world attack scenarios [38]. This leads to poor generalization when deployed in operational environments. The lack of standardized benchmarking datasets across different cybersecurity domains further complicates performance evaluation and comparison.

8.3 Explainability and Transparency

Explainability remains a major challenge, particularly for deep learning and transformer-based cybersecurity models. While these models provide high detection accuracy, their decision-making processes are often opaque, making it difficult for security analysts to interpret results and validate predictions.

The lack of transparency reduces trust in AI-driven systems and limits their adoption in critical sectors such as healthcare, finance, and government infrastructure [39]. Although Explainable AI (XAI) techniques such as SHAP and LIME have been proposed, they often provide approximate or post-hoc explanations that may not fully reflect the internal reasoning of complex models. Achieving a balance between accuracy and interpretability remains a key research challenge.

8.4 Privacy and Regulatory Compliance

Cybersecurity systems frequently process sensitive and confidential data, including user behavior logs, network traffic, and organizational security events. This raises significant concerns regarding data privacy and regulatory compliance. Regulations such as GDPR, HIPAA, and national cybersecurity laws impose strict constraints on how data can be collected, stored, and processed [40].

AI models, particularly those requiring centralized data aggregation, may conflict with privacy requirements. Federated learning offers a promising solution by enabling decentralized training; however, it is still vulnerable to inference attacks and model leakage. Ensuring privacy-preserving AI while maintaining high performance remains an open problem in cybersecurity research.

8.5 Ethical Concerns

The use of AI in cybersecurity introduces several ethical challenges, particularly related to dual-use technologies. Generative AI models, for instance, can be used both for defensive purposes (e.g., synthetic data generation, attack simulation) and malicious activities (e.g., phishing email generation, malware creation, deepfake attacks).

Other ethical concerns include bias in training data, unfair decision-making in risk scoring systems, and the potential misuse of autonomous cyber defense systems. Additionally, the deployment of fully automated security systems raises questions about accountability and responsibility in the event of system failure or misclassification. Establishing ethical guidelines and governance frameworks is essential for responsible AI deployment in cybersecurity.

8.6 Human-AI Collaboration Challenges

Effective cybersecurity operations require seamless collaboration between human analysts and AI systems. However, current AI models often fail to align with human decision-making processes, leading to issues such as alert fatigue, over-reliance on automation, or distrust in AI recommendations.

SOCs frequently generate large volumes of alerts, many of which may be false positives, causing cognitive overload for analysts. Additionally, insufficient interpretability of AI models limits their usefulness in assisting human decision-making. Designing human-centered AI systems that enhance

analyst efficiency while maintaining transparency and control remains a major challenge.

The challenges discussed in this section highlight the complexity of integrating AI into cybersecurity environments. Issues such as adversarial attacks, data limitations, lack of explainability, privacy concerns, ethical risks, and human–AI collaboration barriers must be addressed to fully realize the potential of AI-driven cybersecurity systems. Addressing these challenges requires interdisciplinary research combining AI, cybersecurity engineering, human–computer interaction, and regulatory science. These open issues also motivate the development of hybrid AI–fuzzy decision-making frameworks and next-generation adaptive cybersecurity systems.

9. Proposed AI–Fuzzy Cybersecurity Decision-Making Framework

To address the limitations of existing AI-based cybersecurity solutions, this study proposes an integrated AI–Fuzzy Cybersecurity Decision-Making Framework. The framework is designed to combine the predictive capabilities of AI models with the uncertainty-handling capability of fuzzy logic and the structured reasoning strength of MCDM techniques. The proposed model aims to support intelligent, adaptive, and transparent cybersecurity decision-making in dynamic and high-risk environments.

9.1 Design Objectives

The primary objective of the proposed framework is to enhance cybersecurity decision-making by integrating AI-driven analytics with fuzzy logic and MCDM techniques. The specific design objectives include:

- To enable real-time detection and classification of cyber threats using AI models.
- To handle uncertainty, ambiguity, and incomplete information using fuzzy logic.
- To support multi-criteria risk evaluation and prioritization using MCDM techniques such as AHP and TOPSIS.
- To improve explainability and transparency in cybersecurity decision processes.
- To enable adaptive and continuous learning for evolving cyber threat landscapes.
- To support scalable deployment across cloud, edge, and IoT environments.

9.2 Framework Architecture

The proposed AI–Fuzzy cybersecurity decision-making framework consists of seven interconnected layers, each performing a distinct function in the cybersecurity pipeline.

- **Data Collection Layer:** This layer is responsible for collecting raw cybersecurity data from multiple heterogeneous sources, including network traffic logs, system logs, application logs, intrusion detection systems, firewall logs, IoT device data, and threat intelligence feeds. The data collected may be structured, semi-structured, or unstructured, requiring preprocessing before further analysis.
- **AI Analytics Layer:** The AI analytics layer processes the collected data using machine learning, deep learning, transformer-based models, and anomaly detection algorithms. This layer is responsible for identifying patterns, detecting anomalies, classifying threats, and predicting potential cyberattacks. It serves as the core intelligence engine of the framework.
- **Threat Intelligence Layer:** This layer aggregates and analyzes external and internal threat intelligence data from sources such as security reports, vulnerability databases, open-source intelligence (OSINT), and dark web monitoring. Natural language processing models and LLMs are used to extract relevant insights and contextualize threat information.
- **Fuzzy Inference Layer:** The fuzzy inference layer handles uncertainty and imprecision in cybersecurity data. It converts numerical and linguistic inputs into fuzzy sets and applies fuzzy rules to evaluate risk levels, threat severity, and system vulnerability. Outputs are generated in linguistic terms such as low, medium, and high risk, enabling human-readable interpretation.
- **Decision-Making Layer:** This layer applies MCDM techniques such as the Analytic Hierarchy Process (AHP) and TOPSIS to prioritize cybersecurity risks and determine optimal mitigation strategies. It integrates outputs from the AI analytics and fuzzy inference layers to evaluate multiple conflicting criteria, including risk severity, attack likelihood, system criticality, and mitigation cost.
- **Cybersecurity Response Layer:** The response layer executes appropriate mitigation actions based on decisions generated by the decision-making layer. These actions may include blocking malicious traffic, isolating affected systems, triggering alerts, applying patches, or activating incident response protocols. This layer ensures real-time and automated threat mitigation.
- **Continuous Learning Layer:** The continuous learning layer enables the framework to adapt over time by incorporating feedback from cybersecurity incidents, analyst inputs, and evolving attack patterns. Reinforcement learning and incremental learning techniques are used to update AI models and improve decision accuracy and robustness.

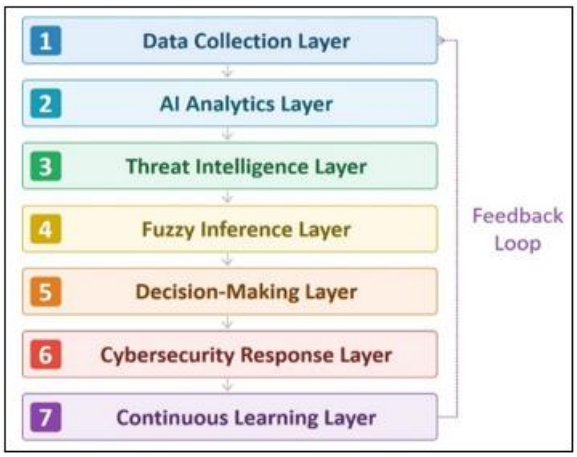

Figure 4: The Framework

9.3 Expected Benefits and Practical Implications

The proposed AI–Fuzzy cybersecurity decision-making framework offers several significant benefits:

- **Improved Detection Accuracy:** Integration of multiple AI models enhances threat detection performance.
- **Uncertainty Handling:** Fuzzy logic effectively manages ambiguity and incomplete information.
- **Better Decision Quality:** MCDM techniques ensure structured and optimized cybersecurity decisions.
- **Enhanced Explainability:** Fuzzy rules and MCDM outputs improve interpretability for human analysts.
- **Adaptive Learning**: Continuous learning mechanisms enable real-time adaptation to emerging threats.
- **Scalability:** Modular architecture supports deployment in cloud, edge, and IoT environments.
- **Reduced Response Time:** Automated decision-making accelerates incident response processes.

From a practical perspective, the framework is applicable to SOCs, enterprise cybersecurity systems, critical infrastructure protection, and cloud security management. It can assist cybersecurity analysts in prioritizing threats, reducing alert fatigue, and improving overall operational efficiency.

10. Discussion

This systematic review demonstrates a major transformation in cybersecurity research from traditional rule-based and signature-based approaches to intelligent, adaptive, and data-driven security architectures. This shift reflects the increasing sophistication of cyber threats, expanding digital infrastructures, and the limitations of conventional methods in detecting unknown attacks. The reviewed studies confirm that AI has become a key enabler of modern cybersecurity by providing predictive, adaptive, and autonomous capabilities for securing complex cyber ecosystems.

A major finding is the widespread adoption of Machine Learning (ML) and Deep Learning (DL) for intrusion detection, malware classification, anomaly detection, phishing detection, and network traffic analysis. Their ability to learn complex patterns from large datasets has significantly improved detection performance. However, challenges remain, including dependence on high-quality labeled data, vulnerability to adversarial attacks, high computational costs, and limited interpretability. These limitations highlight the need for cybersecurity solutions that balance detection accuracy with transparency, trust, and accountability.

Transformer-based architectures and Large Language Models (LLMs) have emerged as significant advances in cybersecurity because of their ability to process unstructured textual data. They support threat intelligence, vulnerability assessment, security log analysis, incident reporting, and cyber threat hunting, enabling more context-aware security systems. However, issues such as hallucinations, prompt manipulation, and misinformation require robust validation and governance before large-scale deployment.

The review also identifies the growing role of Generative AI in cybersecurity. It supports synthetic data generation, adversarial simulation, penetration testing, cyber training, and threat intelligence synthesis, helping address data scarcity and improve defensive readiness. At the same time, these technologies can be misused for phishing, malware generation, social engineering, and automated attacks, emphasizing the need for responsible AI governance.

Reinforcement Learning (RL) is gaining importance for autonomous cybersecurity through adaptive intrusion response, intelligent access control, and Security Operations Centers (SOCs). Although RL enables continuous learning in dynamic environments, challenges related to reward design, computational complexity, and unintended autonomous behavior remain. Human-in-the-loop mechanisms are therefore essential to ensure accountability.

Federated Learning (FL) has emerged as a promising privacy-preserving approach for collaborative cybersecurity analytics. It enables organizations to jointly develop intelligent models without sharing sensitive data, addressing privacy and regulatory concerns. However, issues such as communication overhead, model poisoning, trust management, and heterogeneous data continue to limit its large-scale adoption.

Another important finding is the integration of fuzzy logic and Multi-Criteria Decision-Making (MCDM) methods into AI-driven cybersecurity frameworks. Because cybersecurity decisions involve uncertainty and subjective expert judgment, fuzzy logic effectively models linguistic uncertainty, while techniques such as AHP, TOPSIS, VIKOR, and DEMATEL support systematic evaluation of security alternatives. Their integration with AI improves decision transparency, interpretability, and explainability.

Overall, cybersecurity research is increasingly adopting hybrid intelligence frameworks that combine machine learning, deep learning, fuzzy reasoning, symbolic knowledge, optimization, and decision-support techniques. Examples such as neuro-fuzzy and neuro-symbolic systems demonstrate that integrating multiple methodologies enhances predictive performance while improving explainability, resulting in more reliable and effective cybersecurity decision-making.

11. Conclusion

This paper presented a comprehensive systematic review of emerging AI models in cybersecurity, with a particular focus on their applications, decision-making capabilities, challenges, and future research directions. The study systematically analyzed a wide spectrum of AI paradigms, including machine learning, deep learning, transformer and LLMs, generative AI, reinforcement learning, federated learning, explainable AI, and hybrid neuro-symbolic systems. In addition, the role of fuzzy logic and MCDM techniques was examined to highlight their significance in handling uncertainty and supporting structured cybersecurity decision processes.

The review demonstrates that AI has become a fundamental enabler of modern cybersecurity systems, significantly enhancing capabilities in intrusion detection, malware analysis, phishing detection, threat intelligence, vulnerability assessment, cloud and IoT security, digital forensics, and critical infrastructure protection. Among the examined approaches, deep learning and transformer-based models show strong performance in complex detection and analysis tasks, while traditional machine learning methods remain relevant due to their interpretability and computational efficiency. However, each AI paradigm presents inherent limitations, particularly in terms of explainability, scalability, robustness, and susceptibility to adversarial attacks.

A key contribution of this study is the integration of AI techniques with fuzzy logic and MCDM-based decision-making frameworks. This integration enables more reliable cybersecurity decision support by addressing uncertainty, ambiguity, and multi-criteria trade-offs in dynamic environments. The proposed AI–Fuzzy cybersecurity decision-making framework provides a unified architecture that combines AI analytics, threat intelligence, fuzzy inference, and structured decision-making to support adaptive and intelligent cyber defense mechanisms. The review further identifies critical challenges that must be addressed to advance the field, including adversarial AI attacks, data scarcity and quality issues, lack of transparency, privacy concerns, ethical risks, and limitations in human–AI collaboration. These challenges highlight the need for more robust, explainable, and privacy-preserving cybersecurity solutions capable of operating effectively in real-world scenarios.

Finally, the study outlines several promising future research directions, including explainable and trustworthy AI, generative AI for cyber defense, autonomous SOCs, federated learning-based privacy-preserving systems, quantum-AI security integration, human-centered cybersecurity design, and intelligent AI–fuzzy hybrid platforms. These directions collectively point toward the development of next-generation cybersecurity systems that are intelligent, adaptive, explainable, and resilient. The convergence of AI, fuzzy logic, and advanced decision-making techniques represents a significant paradigm shift in cybersecurity research and practice. The findings of this study provide a structured foundation for researchers and practitioners to develop more effective, transparent, and scalable cybersecurity solutions capable of addressing the evolving complexity of modern cyber threats.

References

[1] M. A. Ferrag, L. Maglaras, A. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," Journal of Information Security and Applications, vol. 50, p. 102419, 2020.
[2] Y. Li, "Deep reinforcement learning: An overview," arXiv preprint arXiv:1701.07274, 2017.
[3] R. A. Khan and M. W. Khan, “Cyber Security’s Influence on Smart Cities: Challenges and Solutions,” in Contemporary Innovations in Engineering and Management, AIP Publishing, vol. 2821, pp. 040033-1-12, 2023.
[4] R. Gupta, A. Sharma, and M. Singh, "Generative AI for cybersecurity: Applications and challenges," IEEE Access, vol. 12, pp. 45012–45035, 2024.
[5] S. K. Dwivedi, J. Yadav, S. A. Ansar, M. W. Khan, D. Pandey and R. A. Khan, "A Novel Paradigm: Cloud-Fog Integrated IoT Approach," 2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, pp. 1-5, 2022.
[6] T. Li, Z. Shi, and X. Li, "A survey on reinforcement learning for cybersecurity," IEEE Transactions on Network and Service Management, vol. 20, no. 3, pp. 2431–2451, 2023.
[7] P. Shrivastava, A. Jamal, and K. Bhardwaj, "Neuro-symbolic AI for cybersecurity decision support: A systematic review," Computers & Security, vol. 138, p. 103677, 2024.
[8] D. Pandey, N. Singh, V. Singh, and M. W. Khan, “Paradigms of Smart Education with IoT Approach,” in Internet of Things and Its Applications, S. N. Mohanty, J. M. Chatterjee, and S. Satpathy, Eds. Cham, Switzerland: Springer, 2022.
[9] Al-Nawasrah, A. I. Awad, A. Mousa, and M. Hammoudeh, "Transformer-based natural language processing for cyber threat intelligence extraction," Future Generation Computer Systems, vol. 141, pp. 120–131, 2023.
[10] Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, p. e4150, 2021.
[11] J. Saxe and K. Berlin, "Deep neural network based malware detection using two dimensional binary program features," in Proc. 10th Int. Conf. Malicious and Unwanted Software (MALWARE), 2015, pp. 11–20.
[12] S. Dutta, D. Ghosh, S. Banerjee, and S. Dey, "A review of machine learning and deep learning approaches for phishing detection," Journal of Cybersecurity and Privacy, vol. 2, no. 3, pp. 606–625, 2022.
[13] C. Liao, K. Yuan, X. Luo, J. Wang, F. Monrose, and G. Wang, "Acing the IOC game: Toward automatic discovery and analysis of open-source cyber threat intelligence," in Proc. ACM SIGSAC Conf. Computer and Communications Security, 2016, pp. 755–766.
[14] A. Mishra, M. H. Khan, W. Khan, M. Z. Khan, and N. K. Srivastava, “A Comparative Study on Data Mining Approach Using Machine Learning Techniques: Prediction Perspective,” in A Compendium of Critical Factors for Success, Pervasive Healthcare, EAI/Springer Innovations in Communication and Computing. Cham, Springer, pp. 153–165, 2022.
[15] Abutaleb and A. Fahad, "Cloud security using artificial intelligence: Techniques and challenges," IEEE Access, vol. 11, pp. 61032–61050, 2023.
[16] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, 2016.

[17] N. Al-Dhaqm, S. Razak, R. A. Ikuesan, V. R. Kebande, and K. Siddique, "A review of mobile forensic investigation process models," IEEE Access, vol. 8, pp. 173359–173375, 2020.
[18] Zhu, A. Joseph, and S. Sastry, "A taxonomy of cyber-attacks on SCADA systems," in Proc. IEEE Int. Conf. Internet of Things (iThings) and IEEE Green Computing and Communications, 2011, pp. 380–388.
[19] E. Triantaphyllou, Multi-Criteria Decision Making Methods: A Comparative Study. Dordrecht: Kluwer Academic Publishers, 2000.
[20] S. A. Ansar, S. P. Srivastava, J. Yadav, M. W. Khan, A. Yadav, and R. A. Khan, "Estimation of Software Risks through CVSS: A Design Phase Perspective," Turkish Online Journal of Qualitative Inquiry, vol. 12, no. 4, pp. 894–901, 2021.
[21] L. Hwang and K. Yoon, Multiple Attribute Decision Making: Methods and Applications. Berlin: Springer-Verlag, 1981.
[22] L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, no. 3, pp. 338–353, 1965.
[23] M. Chandrashekhar and K. Raghuveer, "Confederation of fuzzy control theory and FPGA implementation for efficient intrusion detection and prevention in digital network communication," in Proc. IEEE Int. Conf. Communication and Signal Processing, 2013, pp. 107–113.
[24] S. Ganapathy, P. Yogesh, and A. Kannan, "Intelligent agent-based intrusion detection system using enhanced multiclass SVM," International Journal of Distributed Sensor Networks, vol. 8, no. 9, p. 659483, 2012.
[25] S. Opricovic and G.-H. Tzeng, "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS," European Journal of Operational Research, vol. 156, no. 2, pp. 445–455, 2004.
[26] B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," Pattern Recognition, vol. 84, pp. 317–331, 2018.
[27] Shafahi et al., "Poison frogs! Targeted clean-label poisoning attacks on neural networks," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2018, pp. 6106–6116.
[28] G. Malgieri and G. Comande, "Why a right to legibility of automated decision-making exists in the general data protection regulation," International Data Privacy Law, vol. 7, no. 4, pp. 243–265, 2017.
[29] K. Bonawitz et al., "Towards federated learning at scale: A system design," in Proc. Machine Learning and Systems (MLSys), 2019.
[30] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.
[31] A. Attaallah, K. Al-Sulbi, A. Alasiry, M. Marzougui, M. W. Khan, M. Faizan, A. Agrawal, and D. Pandey, "Security Test Case Prioritization through Ant Colony Optimization Algorithm," Computer Systems Science and Engineering, vol. 47, no. 3, pp. 3165–3195, 2023.
[32] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," IEEE Access, vol. 7, pp. 41525–41550, 2019.
[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
[34] Goodfellow et al., "Generative adversarial nets," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2014, pp. 2672–2680.
[35] Abdulhammed, H. Musafer, A. Alessa, M. Faezipour, and A. Abuzneid, "Features dimensionality reduction approaches for machine learning based network intrusion detection systems," Electronics, vol. 8, no. 3, p. 322, 2019.
[36] M. Shabbir, R. K. Yadav, M. W. Khan, H. Singh, "Empirical Analysis of Computationally Intelligent Technique for Software Risk Prediction." Journal of Cybersecurity and Information Management, vol. Volume 17, no. Issue 2, pp. 200-209, 2026.
[37] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.
[38] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you: Explaining the predictions of any classifier," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[39] M. Schuld, I. Sinayskiy, and F. Petruccione, "An introduction to quantum machine learning," Contemporary Physics, vol. 56, no. 2, pp. 172–185, 2015.
[40] T. D. Nguyen and T. T. Reddi, "Reinforcement learning for cyber operations: A survey," IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1945–1975, 2021.