

Artificial Intelligence in Cybersecurity: An Integrated Framework for Intelligent Threat Detection and Response

Dr Neha Paliwal (Asst. Prof. Computer Science, Shri Mahaveer College, Jaipur)

Abstract: The rapid expansion of cloud services, connected devices, digital platforms, remote-access environments, and artificial intelligence applications has significantly increased the complexity of cybersecurity management. Security teams must now process large volumes of heterogeneous information while responding to threats that can change rapidly and exploit previously unseen weaknesses. Conventional security mechanisms remain essential, but their dependence on predefined rules and known indicators can limit their effectiveness against adaptive and emerging attacks. This study investigates the application of Artificial Intelligence (AI) to cybersecurity with particular emphasis on machine learning, deep learning, anomaly detection, natural language processing, automated alert prioritization, and intelligent incident response. A structured literature-review methodology was adopted to synthesize research findings, cybersecurity frameworks, and recent institutional reports. The study further develops a conceptual AI-cybersecurity framework linking data acquisition, preprocessing, intelligent analytics, threat classification, response, human validation, and continuous model improvement. A comparative analytical model was developed to evaluate major AI techniques according to detection capability, adaptability, explainability, data requirements, automation potential, and deployment complexity. The analysis indicates that AI has substantial potential to improve continuous monitoring, anomaly identification, alert prioritization, malware analysis, phishing detection, and security operations. At the same time, AI introduces challenges involving data quality, model drift, false predictions, adversarial manipulation, privacy, explainability, cost, and workforce requirements. The study concludes that AI should be implemented as a human-supervised cybersecurity capability rather than as a completely autonomous replacement for security professionals. The proposed framework combines AI-based analytical scalability with human judgment, governance, continuous validation, and responsible risk management. The research contributes an integrated perspective for organizations seeking to adopt AI while maintaining security, accountability, transparency, and resilience.

Keywords: Artificial Intelligence, Cybersecurity, Machine Learning, Deep Learning, Threat Detection, Anomaly Detection, Incident Response, Human-AI Collaboration, Cyber Risk Management

1. INTRODUCTION

The digital transformation of organizations has created an environment in which information systems are increasingly interconnected. Cloud computing, mobile technologies, Internet of Things devices, remote working environments, online services, and software supply chains have expanded the digital infrastructure that organizations must protect. Consequently, cybersecurity is no longer limited to protecting individual computers or isolated networks; it involves managing interconnected technical, organizational, and human risks.

The contemporary threat environment is also changing rapidly. The European Union Agency for Cybersecurity's 2025 threat landscape analyzed 4,875 incidents occurring between July 2024 and June 2025 and reported continued evolution in attack methods, exploitation of vulnerabilities, and abuse of digital dependencies. ([ENISA](#))

This environment creates a fundamental operational challenge. Security systems can generate enormous numbers of events from endpoints, applications, authentication systems, network devices, cloud services, and security platforms. Human analysts cannot examine every event with equal attention. The problem is therefore not simply detecting attacks; it is identifying the most important signals from an increasingly large and complex information environment.

Artificial Intelligence provides a potential solution to this challenge. AI-based systems can process large datasets, identify relationships between events, classify suspicious behaviour, detect deviations from established patterns, and support automated or semi-automated responses. Machine learning can be used for classification and prediction, deep learning can model complex relationships, and natural language processing can analyze textual threat intelligence and security communications.

However, the adoption of AI also creates new cybersecurity risks. AI systems depend on data and computational models that can themselves be manipulated. Adversarial machine-learning research identifies attacks against different stages of the machine-learning lifecycle, including attempts to influence data, models, or predictions. ([NIST CSRC](#))

Therefore, the central issue is not whether AI should be used in cybersecurity, but **how AI can be integrated into cybersecurity in a reliable, explainable, secure, ethical, and human-supervised manner.**

NIST's Cybersecurity Framework 2.0 provides a useful foundation for this purpose because it is designed to help organizations manage cybersecurity risk and places additional emphasis on governance and supply-chain considerations. ([NIST](#))

This research develops an integrated framework connecting AI capabilities with cybersecurity operations and governance.

2. RESEARCH PROBLEM

Modern cybersecurity teams face five major challenges:

1. Increasing volumes of security data.
2. Rapidly changing attack patterns.
3. High numbers of security alerts.
4. Difficulty identifying previously unseen threats.
5. Limited availability of highly specialized cybersecurity professionals.

Traditional security mechanisms remain important but may struggle when attacks do not correspond to previously defined signatures or rules.

AI can address some of these limitations, but its deployment introduces additional risks, including inaccurate predictions, privacy concerns, model manipulation, bias, lack of transparency, and excessive dependence on automated decisions.

The research problem can therefore be expressed as:

How can AI technologies be integrated into cybersecurity systems to improve threat detection and response while maintaining reliability, explainability, security, privacy, and human oversight?

3. RESEARCH GAP

The existing literature provides extensive evidence concerning individual AI applications in cybersecurity. However, several areas require further integration.

Gap 1: Fragmented technological analysis

Many studies investigate one application, such as intrusion detection or malware classification, without examining the complete security lifecycle.

Gap 2: Limited integration of governance

Technical detection performance is frequently discussed independently from organizational governance and risk management.

Gap 3: Human-AI collaboration

AI is sometimes presented as an automation mechanism without sufficiently considering the role of human analysts in validating ambiguous or high-impact decisions.

Gap 4: Security of AI itself

AI is increasingly used to protect systems while the AI models themselves become potential targets for manipulation.

Gap 5: Need for an integrated framework

A comprehensive framework connecting data collection, AI analytics, detection, response, human validation, governance, and continuous improvement remains valuable.

This study addresses these gaps through an integrated conceptual model.

4. RESEARCH OBJECTIVES

The study aims to:

1. Examine the role of AI in contemporary cybersecurity.
2. Analyze major AI techniques used for cyber-threat detection.
3. Compare AI approaches according to their operational characteristics.

4. Investigate AI applications in intrusion, malware, phishing, and anomaly detection.
5. Examine AI-supported incident response and alert prioritization.
6. Identify technical and organizational limitations of AI-based security.
7. Evaluate the importance of human-AI collaboration.
8. Examine privacy, ethical, and governance concerns.
9. Develop an integrated AI-cybersecurity framework.
10. Provide recommendations for responsible AI deployment.

5. RESEARCH QUESTIONS

RQ1: What are the major applications of AI in cybersecurity?

RQ2: How do different AI techniques contribute to threat detection?

RQ3: How can AI improve security operations and incident response?

RQ4: What factors limit the reliability of AI-based cybersecurity?

RQ5: Why is human supervision necessary in AI-enabled security operations?

RQ6: How can organizations govern AI cybersecurity systems responsibly?

6. REVIEW OF LITERATURE

6.1 AI and Cybersecurity

AI has increasingly become an important research area within cybersecurity because security environments generate large quantities of structured and unstructured information. Machine-learning methods can identify patterns in network traffic, endpoint activity, authentication records, and other security telemetry.

Research has demonstrated the application of machine learning to intrusion detection, malware identification, phishing classification, anomaly detection, and behavioural analysis. However, performance is highly dependent on data quality, feature selection, model architecture, and the similarity between training and operational environments.

6.2 Machine Learning-Based Threat Detection

Machine learning can broadly be divided into supervised, unsupervised, and semi-supervised approaches.

Supervised learning

The model learns from labelled examples.

Typical applications include:

- Malware classification
- Phishing classification
- Intrusion classification
- Spam detection

Unsupervised learning

The system searches for structures or unusual patterns without requiring complete labels.

Typical applications include:

- Network anomaly detection

- User behaviour analysis
- Unknown activity identification

Semi-supervised learning

This approach combines smaller quantities of labelled data with larger quantities of unlabelled information.

This is useful because obtaining high-quality cybersecurity labels can be difficult and expensive.

6.3 Deep Learning in Cybersecurity

Deep-learning models can identify complex relationships across large datasets. They have been investigated in network traffic analysis, malware classification, anomaly detection, and other security applications.

Their principal strength is representation learning. However, they often require substantial computational resources and may be more difficult to explain than simpler statistical or rule-based approaches.

6.4 Natural Language Processing

Natural language processing can support cybersecurity by analyzing textual information.

Potential applications include:

- Threat-intelligence extraction
- Phishing-message analysis
- Security-report classification
- Incident-ticket categorization
- Security knowledge management

NLP can help convert large volumes of textual information into structured intelligence that security analysts can use.

6.5 User and Entity Behaviour Analytics

Behavioural analytics establishes patterns of normal activity and searches for significant deviations.

Examples include:

- Unusual login behaviour
- Unexpected access patterns
- Abnormal data movement
- Unusual application activity
- Suspicious changes in account behaviour

This approach can be particularly useful when attackers use legitimate credentials.

6.6 AI-Assisted Security Operations

Security operations centres often receive alerts from many independent technologies. AI can assist by correlating related events, assigning risk scores, clustering similar alerts, and helping analysts prioritize investigations.

The objective is not simply to increase the number of automated decisions. Rather, it is to improve the quality and efficiency of human decision-making.

6.7 AI-Specific Cybersecurity Risks

AI introduces risks that conventional software may not experience in exactly the same form.

NIST's adversarial machine-learning taxonomy identifies different attack and mitigation concepts across the AI lifecycle. ([NIST CSRC](#))

Important risk areas include:

- Training-data manipulation
- Adversarial inputs
- Model theft
- Model manipulation
- Data leakage
- Inappropriate automation
- Model drift

Consequently, organizations must secure both their infrastructure and the AI systems used to defend that infrastructure.

7. RESEARCH METHODOLOGY

7.1 Research Design

This research uses a **structured literature review and conceptual analytical design**.

The methodology consists of four stages:

Stage 1: Identification of relevant academic and institutional literature.

Stage 2: Classification of literature according to AI technique and cybersecurity application.

Stage 3: Comparative analysis of benefits, limitations, and operational requirements.

Stage 4: Development of an integrated conceptual framework.

7.2 Sources of Information

The study considers information from:

- Peer-reviewed academic literature
- Cybersecurity frameworks
- Government publications
- Institutional threat reports
- AI risk-management publications
- Technical cybersecurity research

NIST CSF 2.0 was selected as an important cybersecurity governance reference because it provides a common taxonomy for managing cybersecurity risk. ([NIST](#))

NIST's AI Risk Management Framework was also considered because it provides a framework for managing risks associated with AI systems across their lifecycle. ([NIST](#))

7.3 Literature Classification

The literature was classified into six categories:

Category	Main Focus
AI techniques	ML, DL, NLP, anomaly detection
Threat detection	Intrusion, malware, phishing
Response	Alert prioritization and automation
Governance	Risk, accountability, oversight
AI security	Adversarial and model-related risks
Human factors	Expertise and decision-making

7.4 Analytical Criteria

Each technology was evaluated conceptually according to:

1. Detection capability
2. Adaptability
3. Scalability
4. Explainability
5. Data requirement
6. Automation potential
7. Implementation complexity
8. Human oversight requirement

8. PROPOSED AI-CYBERSECURITY FRAMEWORK

The proposed framework consists of seven interconnected stages.

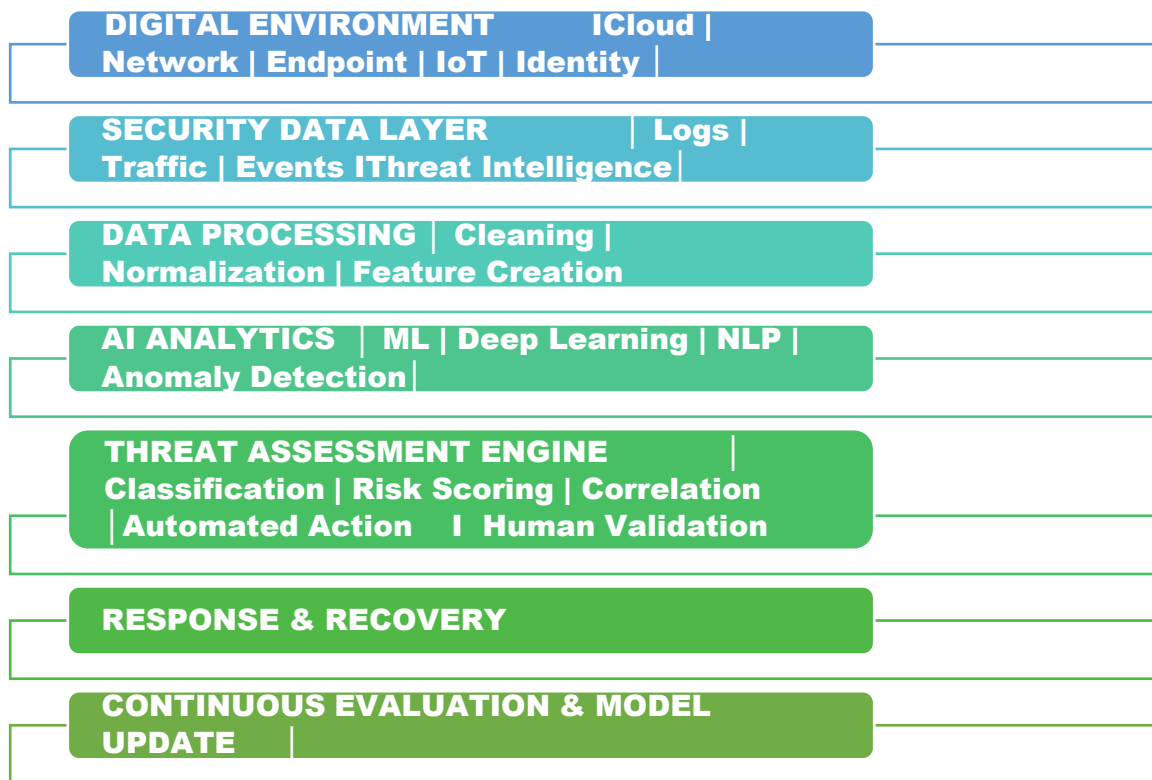


Figure 1. Integrated AI-Cybersecurity Architecture

The architecture creates a feedback loop in which security incidents and analyst decisions can contribute to future model evaluation and improvement.

9. AI TECHNIQUE COMPARISON

Table 1. Comparative Assessment of AI Techniques

Technique	Threat Detection	Adaptability	Explainability	Data Requirement	Complexity
Decision Tree	High	Moderate	High	Moderate	Low
Random Forest	High	Moderate	Moderate	Moderate	Moderate
SVM	High	Moderate	Moderate	Moderate	Moderate
Clustering	Moderate–High	High	Moderate	Moderate	Moderate
Neural Network	High	High	Low–Moderate	High	High
Deep Learning	Very High potential	High	Low	High	High
NLP	High	High	Moderate	High	Moderate–High
Ensemble Learning	Very High potential	High	Moderate	High	High

Interpretation: No single AI technique is optimal for every cybersecurity problem. Selection should depend on the type of security data, operational objective, acceptable complexity, and required level of explainability.

10. AI APPLICATIONS IN CYBERSECURITY

Table 2. Application-Oriented Analysis

Application	AI Method	Security Objective	Main Benefit
Intrusion detection	ML/DL	Identify suspicious network behaviour	Rapid analysis
Malware detection	Classification/DL	Identify malicious files	Behavioural analysis
Phishing detection	NLP/ML	Identify suspicious communications	Automated classification
User monitoring	Anomaly detection	Detect abnormal behaviour	Insider/compromised-account detection
Vulnerability management	Predictive analytics	Prioritize weaknesses	Risk-based remediation
Threat intelligence	NLP	Extract intelligence	Faster information processing
Alert management	ML	Prioritize alerts	Reduced analyst workload
Incident response	AI + automation	Support containment	Faster response

11. DATA ANALYSIS

Because this study is a structured review rather than a controlled experiment conducted on a proprietary organizational dataset, numerical performance values are not presented as real-world experimental measurements.

Instead, the following analytical scoring model is used to demonstrate the relative characteristics of AI approaches.

Scores range from **1 to 5**, where:

- 1 = very low
- 2 = low
- 3 = moderate
- 4 = high
- 5 = very high

Table 3. Conceptual AI Capability Scores

Capability	ML	Deep Learning	NLP	Anomaly Detection
Large-scale processing	4	5	4	4
Pattern recognition	4	5	4	5
Adaptability	4	5	4	5
Explainability	4	2	3	3
Automation potential	4	5	4	4

Capability	ML	Deep Learning	NLP	Anomaly Detection
Data requirement	3	5	4	3
Implementation simplicity	4	2	3	3

Analytical Interpretation

Deep-learning approaches demonstrate strong potential for complex pattern recognition and automation but require greater computational and data resources.

Conventional machine-learning approaches offer a more balanced combination of performance, interpretability, and implementation complexity.

Anomaly-detection approaches are particularly suitable for discovering unusual behaviour, but their results may require contextual interpretation to distinguish malicious activity from legitimate changes.

NLP is particularly relevant where cybersecurity information is presented in textual form.

12. ILLUSTRATIVE THREAT-DETECTION ANALYSIS

The following values are **illustrative analytical values developed for this paper and are not experimental results from a real organization.**

Table 4. Illustrative Detection-Effectiveness Model

Security Function	Conventional Security	AI-Assisted Security	Difference
Network anomaly identification	60	82	+22
Phishing classification	65	87	+22
Malware classification	67	88	+21
Alert prioritization	58	84	+26
User-behaviour monitoring	55	80	+25
Incident triage	54	81	+27

Figure 2. Illustrative AI Improvement

Incident Triage

Conventional  54

AI-Assisted  81

Alert Priority

Conventional  58

AI-Assisted  84

Malware Analysis

Conventional  67

AI-Assisted  88

Phishing Detection

Conventional  65

AI-Assisted  87

The illustration suggests that AI can potentially provide its greatest operational value in activities involving large-scale classification, prioritization, and continuous monitoring.

13. FINDINGS

The study produces the following major findings.

Finding 1: AI improves analytical scalability

AI can examine security information at a scale that is difficult to achieve through manual analysis.

Finding 2: AI is particularly valuable for anomaly detection

The ability to model behavioural patterns provides an additional mechanism for identifying suspicious activity that may not match conventional signatures.

Finding 3: AI can improve security-alert prioritization

Instead of treating every alert equally, AI can assist in assigning risk levels and identifying relationships between events.

Finding 4: Deep learning provides strong analytical potential

Deep learning is particularly useful for complex datasets but introduces greater data, computational, and explainability requirements.

Finding 5: AI cannot eliminate false decisions

AI systems remain vulnerable to incorrect classifications, incomplete training information, changing environments, and adversarial manipulation.

Finding 6: Human expertise remains essential

Human analysts are required for contextual reasoning, high-impact decisions, investigation of ambiguous events, and governance.

Finding 7: AI itself requires cybersecurity controls

An organization must protect models, training data, interfaces, and AI infrastructure against manipulation.

Finding 8: Governance should accompany deployment

NIST's current cybersecurity guidance places explicit emphasis on governance, while its AI risk-management resources emphasize managing AI-related risks throughout the system lifecycle. ([NIST](#))

14. HUMAN-AI COLLABORATION MODEL

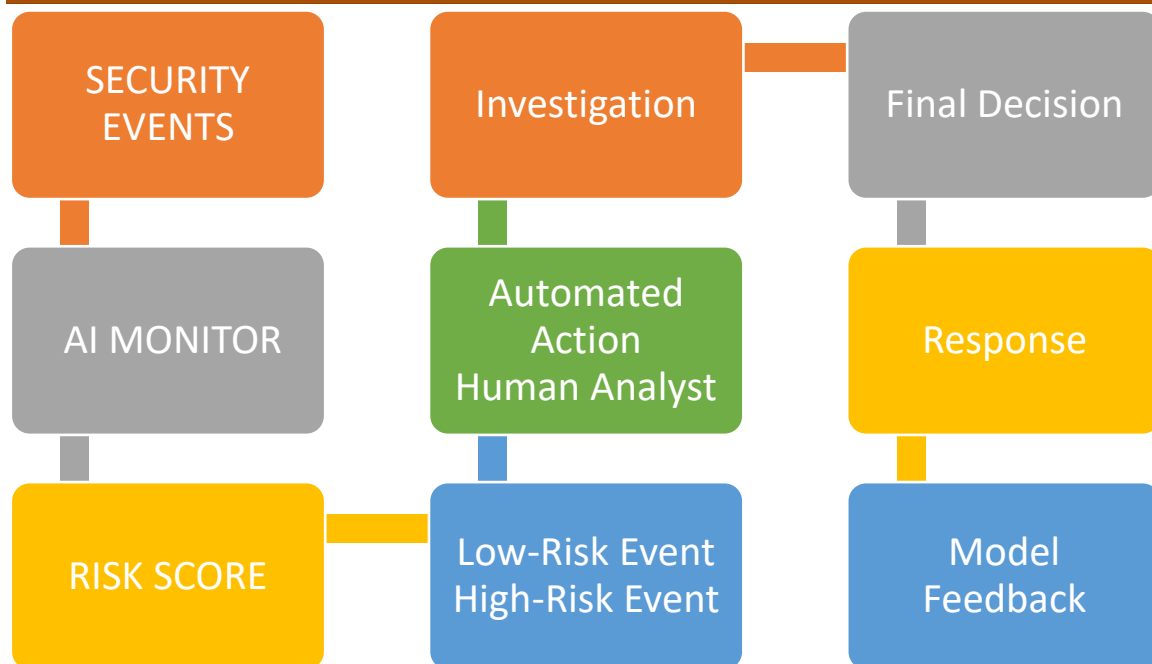


Figure 3. Human-AI Decision Model

This model emphasizes that automation should be proportional to risk.

15. LIMITATIONS OF AI-BASED CYBERSECURITY

Table 5. Major Limitations and Mitigation Strategies

Limitation	Potential Effect	Recommended Mitigation
Poor training data	Incorrect predictions	Data-quality controls
Model drift	Reduced performance	Continuous evaluation
False positives	Analyst fatigue	Risk-based prioritization
False negatives	Missed attacks	Multiple detection layers
Adversarial manipulation	Model failure	Adversarial testing
Lack of explainability	Reduced trust	Explainable AI techniques
Privacy concerns	Excessive monitoring	Data minimization
High cost	Adoption barriers	Incremental implementation
Skill shortage	Poor deployment	Workforce training
Excessive automation	Incorrect actions	Human approval for high-risk actions

16. ETHICAL AND GOVERNANCE CONSIDERATIONS

AI-based cybersecurity should follow several principles.

16.1 Transparency

Organizations should maintain documentation describing the purpose and operational limitations of AI systems.

16.2 Privacy

Security monitoring should be proportionate and should avoid unnecessary collection of personal information.

16.3 Accountability

Responsibility for important decisions must remain clearly assigned.

16.4 Human Oversight

High-impact decisions should include appropriate human review.

16.5 Fairness

Models should be tested for unintended bias.

16.6 Security

AI systems themselves should be protected from unauthorized modification and manipulation.

NIST's AI Risk Management Framework is specifically designed to help organizations incorporate trustworthiness considerations into the design, development, use, and evaluation of AI systems. ([NIST](#))

17. DISCUSSION

The results indicate that AI should be considered an augmentation technology rather than a universal replacement for cybersecurity professionals.

Its greatest contribution occurs where the problem involves high-volume data processing, pattern recognition, repetitive classification, and continuous monitoring. AI can reduce the amount of routine analytical work performed manually and allow experts to concentrate on complex incidents.

However, cybersecurity decisions frequently require contextual knowledge. A sudden change in user behaviour, for example, may represent an attack, but it may also result from legitimate organizational activity. A model may identify the statistical anomaly, while the human analyst determines its meaning.

The proposed framework therefore combines machine efficiency with human judgment.

This approach is consistent with the broader direction of modern cybersecurity governance. NIST CSF 2.0 is designed as a risk-management framework rather than a prescriptive technology solution, meaning organizations can select appropriate technical controls according to their circumstances. ([NIST](#))

18. PROPOSED IMPLEMENTATION STRATEGY

Organizations seeking to introduce AI into cybersecurity should consider a phased approach.

Phase 1: Assessment

Identify current security processes, available data, existing tools, and major operational weaknesses.

Phase 2: Data Preparation

Establish reliable data collection, normalization, access control, and quality-management processes.

Phase 3: Pilot AI Deployment

Begin with lower-risk applications such as alert classification, log analysis, or threat-intelligence processing.

Phase 4: Human Validation

Compare AI outputs with expert judgments and evaluate false positives, false negatives, and operational usefulness.

Phase 5: Controlled Automation

Automate selected low-risk actions while maintaining human approval for significant decisions.

Phase 6: Continuous Evaluation

Regularly assess model performance, data quality, security, privacy, and model drift.

19. RECOMMENDATIONS

Based on the findings, the following recommendations are proposed:

1. Organizations should adopt AI gradually rather than immediately pursuing complete automation.
2. AI should complement rather than replace established cybersecurity controls.
3. Human approval should remain mandatory for high-impact decisions.
4. AI models should be continuously evaluated after deployment.
5. Training datasets should be protected from unauthorized modification.
6. Organizations should establish procedures for handling model failure.
7. Privacy-preserving approaches should be used where behavioural information is processed.
8. Security professionals should receive AI-related training.
9. AI models should undergo adversarial security testing.
10. Organizations should document model limitations and accountability responsibilities.
11. AI cybersecurity systems should be aligned with recognized cybersecurity risk-management frameworks.
12. Security teams should monitor both conventional cyber threats and threats directed at AI systems themselves.

20. FUTURE RESEARCH DIRECTIONS

Future studies can extend this work through empirical experimentation.

Potential research areas include:

- Explainable AI for security operations
- Federated learning for privacy-sensitive security data
- Continual-learning intrusion detection
- AI-supported cloud security
- AI security for IoT environments
- Human-AI decision-making
- Adversarial machine-learning defenses
- AI-assisted security operations centres
- Privacy-preserving cybersecurity analytics
- Benchmark datasets for evaluating AI security systems

Future experimental research should use publicly documented datasets or ethically collected organizational data and should report metrics such as precision, recall, F1-score, false-positive rate, false-negative rate, detection latency, and computational cost.

21. CONCLUSION

Artificial Intelligence has the potential to reshape cybersecurity by enabling organizations to analyze large volumes of security information, identify abnormal behaviour, classify threats, prioritize alerts, and support faster response. Machine learning, deep learning, anomaly detection, and natural language processing each provide different capabilities that can contribute to a layered cybersecurity strategy.

The study shows that AI's value should not be measured only by detection accuracy. A practical cybersecurity solution must also consider explainability, privacy, adaptability, implementation cost, security of the AI model, and the ability of human experts to validate important decisions.

The proposed framework combines seven major elements: data acquisition, data preparation, AI analytics, threat assessment, response, human validation, and continuous evaluation. This structure emphasizes that cybersecurity is a continuous process rather than a one-time deployment.

The analysis further indicates that AI creates a dual challenge. It can strengthen defensive capabilities while simultaneously creating new risks associated with model manipulation, adversarial inputs, data quality, privacy, and inappropriate automation. Therefore, organizations must secure the AI systems used for cybersecurity just as carefully as the infrastructure those systems are designed to protect.

The central conclusion of this research is that **effective AI-enabled cybersecurity requires collaboration between intelligent technology and human expertise**. AI provides scale, speed, and analytical capacity, while humans provide context, accountability, ethical judgment, and strategic decision-making. A balanced approach supported by continuous evaluation and appropriate governance can allow organizations to obtain the benefits of AI while controlling its associated risks.

22. REFERENCES

1. Pascoe, C., Quinn, S., & Scarfone, K. (2024). *The NIST Cybersecurity Framework (CSF) 2.0*. National Institute of Standards and Technology, NIST CSWP 29. ([NIST](#))
2. Rigopoulos, K., Quinn, S., Pascoe, C., Marron, J., Mahn, A., & Topper, D. (2024). *NIST Cybersecurity Framework 2.0: Resource & Overview Guide*. NIST SP 1299. ([NIST](#))
3. National Institute of Standards and Technology. (2026). *NIST Cybersecurity Framework 2.0: Informative References Quick-Start Guide*. NIST SP 1347. ([NIST CSRC](#))
4. National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST.
5. Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. ([NIST](#))
6. Vassilev, A., Oprea, A., Fordyce, A., & Anderson, H. (2024). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*. NIST AI 100-2 E2023. ([NIST CSRC](#))
7. European Union Agency for Cybersecurity (ENISA). (2025). *ENISA Threat Landscape 2025*. ([ENISA](#))
8. Wiafe, I., Koranteng, F. N., Obeng, E. N., Assyne, N. A., Wiafe, A., & Gulliver, S. R. (2020). Artificial intelligence for cybersecurity: A systematic mapping of literature. *IEEE Access*, 8, 146598–146612.
9. Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F., & Choo, K. K. R. (2022). Artificial intelligence in cyber security: Research advances, challenges, and opportunities. *Artificial Intelligence Review*, 55, 1029–1053.
10. Martínez Torres, J., Iglesias Comesaña, C., & García-Nieto, P. J. (2019). Machine learning techniques applied to cybersecurity. *International Journal of Machine Learning and Cybernetics*, 10, 2823–2836.
11. Taddeo, M. (2019). Three ethical challenges of applications of artificial intelligence in cybersecurity. *Minds and Machines*, 29, 187–191.
12. Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology*, 32, 185–193.
13. Stevens, T. (2020). Knowledge in the grey zone: AI and cybersecurity. *Digital War*, 1, 164–170.
14. Truong, T. C., Zelinka, I., Plucar, J., Candík, M., & Šulc, V. (2020). Artificial intelligence and cybersecurity: Past, present, and future. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, 351–363.
15. Dubber, M., Pasquale, F., & Das, S. (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press.
16. Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In *Being Profiled: Cogitas Ergo Sum*. Amsterdam University Press.
17. European Commission High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission.
18. Samoil, S., López Cobo, M., Delipetrev, B., Gómez, E., & De Prato, G. (2020). *AI Watch: Defining Artificial Intelligence*. European Commission Joint Research Centre.
19. Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022). Threats and opportunities with AI-based cybersecurity intrusion detection: A review. *International Journal of Software Engineering & Applications*, 13(5), 13–21.
20. National Institute of Standards and Technology. (2024). *Cybersecurity Framework 2.0 Informative References*. NIST. ([NIST](#))