

An Analysis of Explainable Artificial Intelligence for Intelligent Cybersecurity Applications

Mr. Raman Kumar ¹

¹ Assistant Professor, Department of Computer Science and Applications, Mandsaur (M.P.), India

ABSTRACT: Proactive cyber defenses, AI-powered threat detection systems, and automated reactions are changing the face of cybersecurity in the modern era. In security-critical applications, however, Explainable Artificial Intelligence (XAI) is in high demand due to the need to improve trust, transparency, and decision-making in light of the fact that traditional AI models are not transparent. Intelligent threat detection and response mechanisms, key cybersecurity applications, the most recent advances in the area of XAI-based security solutions, and the fundamentals of explainable AI are all covered in detail in this survey. Highlighting the most prominent explainability methods including SHAP, LIME, Grad-CAM, and counterfactual explanations, this article delves into their applications in several security domains, including cloud security, IoT security, fraud detection, intrusion detection, phishing, spam, cloud security, and identity and access management. The comparative analysis of recent studies is used to emphasize current accomplishments, problems, and new research areas. The results show that XAI can improve the transparency, trustworthiness and effectiveness of AI-based cybersecurity systems, in addition to highlighting a range of privacy, adversarial robustness, scalability and evaluation challenges that warrant further research to ensure reliable deployment in the real world.

KEYWORDS: Intelligent Cybersecurity, Threat Detection, Incident Response, Explainable AI, Intrusion Detection, Network Security.

1. Introduction

The term "cybersecurity" refers to the steps taken to prevent cyberattacks, illegal access, or data loss on computer systems, networks, programs, and data. Data security, integrity, and availability are ensured by cybersecurity's multi-layer security strategy, which incorporates people, procedures, and technology. Network, application, and cloud security are crucial components of cyber defense. Data must be protected in the cyber environment using a multi-layered security policy to ensure its confidentiality, integrity, and availability [1]. Various security technologies, including firewalls, encryption, and AI-based security, as well as novel security concepts like the Zero Trust model, are employed, with a reliance on human awareness, to mitigate security risks [2][3]. Superior safety measures are required in a world where over 61% of all social and commercial interactions take place online. Issues such as cybersecurity, privacy, reliability, and availability should be addressed in these standards to guarantee interactions that are safe, efficient, and contain no data breaches [4]. Maintaining secure inter-industry and societal connections has led to cybersecurity being the

primary method for combating cybercrime and cyberattacks.

The use of AI and ML seems to mark a significant change towards developing more proactive and resilient cybersecurity methods, thanks to their capacities to learn from large data sets, spot trends, and anticipate potential weaknesses. Not only can these technologies enhance existing security processes, they also open the door to completely new ways of detecting, responding to, and preventing threats. This paper aims to critically examine recent developments in cybersecurity-related artificial intelligence and ML [5].

Cybersecurity solutions powered by AI have surfaced in reaction to these instances, assisting teams in better reducing risks and bolstering overall security. Additionally, cybersecurity researchers have started looking into AI-based solutions, such as Gen AI, computer vision, expert systems, deep learning, generative AI, NLP, fuzzy logic, and ML [6]. They are replacing the traditional approaches like firewalls and signature-based IDS with the objective of increasing the accuracy of cyberdefense systems [7][8]. AI models have

demonstrated significant potential in the field of cybersecurity, however, there is one major problem with them: many AI models are black boxes. They have opaque decision-making processes and are unreliable and unreadable for human operators [9]. Some studies explore AI in cybersecurity but not as much those that delve into the significance of explainability in cybersecurity systems. Although the literature already contains several works that consider performance measures of detection models, few studies focus on the understanding of the decision-making process of detection models and the interpretation of the decisions of detection models to the user [10].

A. Structure of the paper

The structure of the paper is as follows: Section II discusses intelligent cyber threat response and detection. In Section III, the concepts of cybersecurity and Explainable AI are defined and introduced. In Section IV the literature review emphasized as well as the gaps in the literature. The final part of the report Section V covers possible directions for future research.

2. Intelligent Cyber Threat Detection and Response

Intelligent cyber threat detection and response requires ongoing monitoring, quick detection, analysis and mitigation of cyber threats [11]. It uses real-time monitoring, threat intelligence, behavioral analysis and automated security to detect suspicious activity. These features enable enterprises to proactively prepare for attacks, act faster, minimize losses, and fortify their cyber resilience [12].

A. Real-Time Threat Detection

One of the most important elements of contemporary cybersecurity strategies. Conventional techniques such as signature-based detection have proven ineffective in the face of the constantly evolving landscape of cybersecurity threats (Stallings, 2020) [13]. Unlike signature-based detection, which relies on known patterns or signatures to identify threats, zero-day attacks are entirely new and unexpected [14].

B. Anomaly Detection

Anomaly detection, also known as outlier detection, is the process of finding data points that are very different from the rest. Data that is significantly different from the rest is not necessarily incorrect, they are called anomalies. This challenge has a wide range of real-world applications, from spotting health issues to recognizing unhappy consumers, from detecting failures to discovering financial crime [15]. There are usually three different types of outliers:

- Isolated anomalies: An anomaly data point that is different from the rest of the data set is called an isolated

anomaly or point anomaly.

- Contextual anomalies: A contextual anomaly is an observation that is considered typical in one context, but unusual in another.
- Collective anomalies: It is crucial to analyze a series to detect collective anomalies.

C. Automated Incident Response

Manual incident response procedures are laborious and open to mistakes made by humans. After an incident, first responders have to manually handle tremendous volumes of data that can end up making the situation last longer. AIR (Automated Incident Response) is the answer. The detection of incidents, analysis and resolution can be accelerated by self-healing systems and ML [16]. There is a risk of increased operating expenses and downtime due to delayed replies and inaccurate evaluations caused by traditional incident response procedures.

D. Threat Detection and Prediction

The ability of ML to search through data from a multitude of sources, including logs of endpoint activity, logs of user behavior, and network logs, for potential hazards [17]. Attack vectors are predicted when the models identify patterns in past and present data, which can help organizations take proactive action to enhance their security practices.

E. Continuous Monitoring

Real-time monitoring, networks and data to detect security vulnerabilities and threats. This method is not periodic assessment method because it allows monitoring and analyze the data regarding security issues continuously, and this helps to maintain security compliance with security policies and regulations [18]. This proactive strategy can significantly reduce the risk of security concerns causing operational disruptions and damages by taking action in a timely and effective manner.

F. Intelligent cybersecurity applications

Intelligent cybersecurity applications are tailored to support cybersecurity in specific contexts, providing a variety of functions from proactive threat detection to automated incident response, continuous monitoring, and adaptive defense. Key applications include intrusion detection, ransomware protection, network security, cloud security, IoT security, endpoint protection, identity management, and intelligent threat hunting for emerging cyber risks [19][20].

1) Intelligent Intrusion Detection

IDS are a standard component of most security

infrastructures. The IDS is a powerful tool for keeping computers safe from intrusion [21]. A wide range of audit and network traffic monitoring techniques, together with other targeted sources of activity in computer or network systems, are utilized to provide security services.

2) Network Security Model

The goal is to have a message sent from one location to another via an online service. A neutral third party might be tasked with delivering the confidential message to both the sender and the receiver while safeguarding it from any potential adversary. When safeguarding data transmission from an adversary who

may compromise its authenticity, secrecy, or other properties becomes important, security considerations come into play; Figure 1 depicts the paradigm of network security [22]. There are always two parts to every security method:

- A change to the data that is about to be transmitted that is linked to security. To make the communication unintelligible to the adversary, it should be encrypted using a key.
- An encryption key that, when combined with the transformation, encrypts the message before sending it and decrypts it upon receiving it.

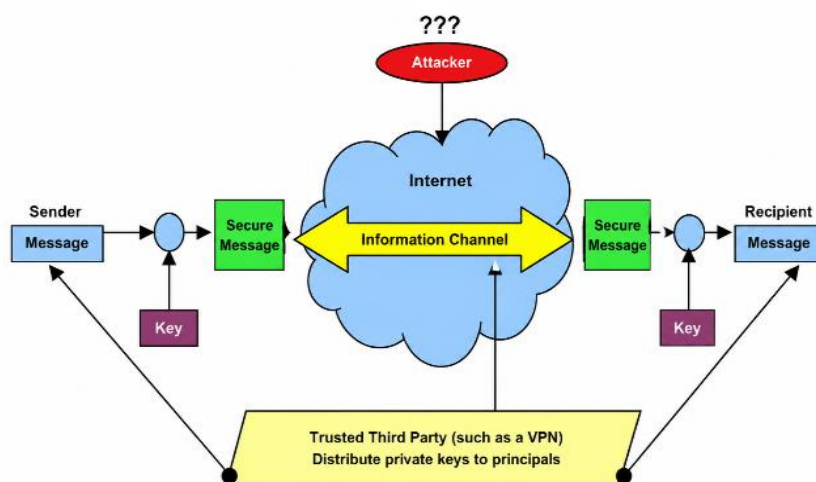


Fig. 1. Model for Network Security

3) Cloud Security

Cloud security encompasses cloud computing and determines an arrangement of guidelines, controls, and innovations that are helpful for defending information and services. The cloud framework is influenced directly or indirectly by threats and attacks [23]. Compromise of cloud resources' integrity, availability, and confidentiality, as well as the services provided at various levels, may generate new security concerns.

4) IoT Security

Secure communication, access control, constant monitoring, threat detection, and vulnerability management are the pillars of IoT security, which shields data, networks, and connected devices from cybercriminals. This ensures that data remains private, undamaged, and reliably runs [24].

5) Identity and Access Management

Integrating "identity management" with "access management," IAM addresses both of these concerns. It incorporates things like identity management and the procedures linked to identification and authorization in businesses [25]. Any living thing, whether human or

not, has the ability to assert its identity. Distinguishing between the human entities of suppliers, customers, partners, and workers is possible in an enterprise setting [26]. Groups, machines, or software programs that do not consist of humans are considered non-human entities.

3. Fundamentals of Explainable Artificial Intelligence and Cybersecurity

A set of methods aimed at making the judgments made by AI models more accessible to humans is collectively known as "explainable artificial intelligence" (XAI). Whereas traditional black-box models conceal the characteristics that influence outcomes and the forecasting process, XAI exposes them. To better detect threats and make educated security decisions, XAI enhances cybersecurity by increasing transparency, trust, accountability, and analyst confidence. When AI is applied in areas with high stakes, where judgments need to be transparent, auditable, and fair, explainability becomes even more important. Ensuring legal and regulatory compliance, promoting trust, improving user adoption, and assisting debugging and model refinement are all important functions it performs.

A. Importance of XAI in Cybersecurity

XAI is more than just about interpretability. In today's age of advanced and personalized cyberattacks, comprehending the means by which AI systems categorize and identify risks is essential for ensuring safety [27]. FP in safety alerts is decreased using XAI. AI can generate a large number of false alarms, which can overwhelm and divert security staff. With XAI's explanations of AI results, security teams may make better use of both human sense and data supplied by AI. Together, strengthen cybersecurity by giving professionals in the field the tools they need to make choices based on data while still keeping tabs on critical security activities. When it comes to cybersecurity, XAI

is crucial for making AI-driven solutions transparent, responsible, and successful. Businesses may be certain that they can lessen the impact of cyber risks, do away with FP, and fortify themselves against hostile assaults. Because cyber threats evolve and there is a need for AI systems with more explainability, XAI is crucial to contemporary cybersecurity frameworks. Uses of XAI in cyber defense [28]. Figure 2 shows that XAI is becoming more important in the battle against various cyberattacks. Here looked at the most cutting-edge XAI-based defensive solutions for various types of cyberattacks.

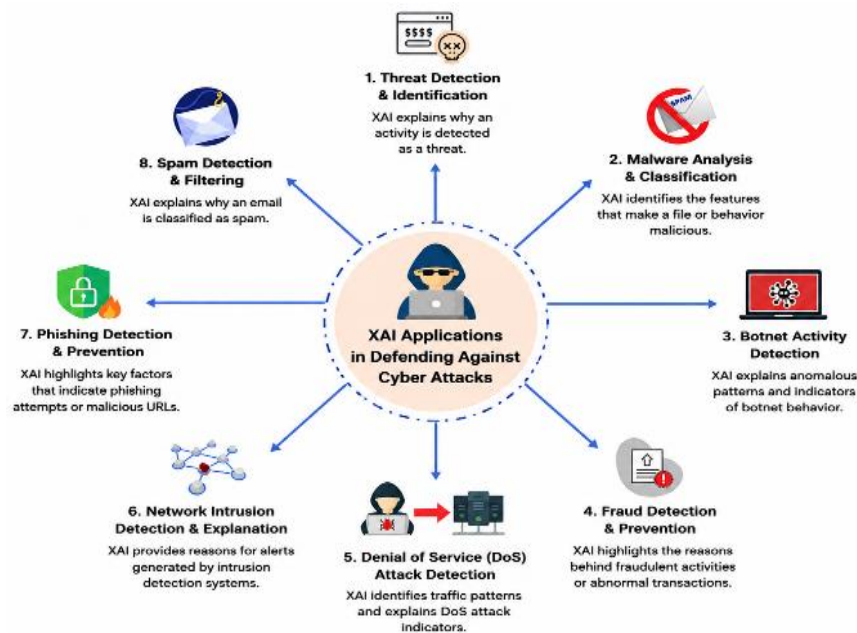


Fig. 2. XAI applications in defending against cyberattacks

1) Malware

Malware is one of the most significant cybersecurity threats on the Internet today; hence, rapid analysis is required to establish effective defenses. These days, there are primarily two kinds of algorithms used to identify malware: It distinguishes between static and dynamic detection [29].

- **Static malware detection:** examines the malicious executable file without executing the program. Instead, decompilation tools are used to obtain the source or assembly code, and the extracted instructions are examined to identify malicious patterns.

- **Dynamic malware detection:** On the other hand, it runs the malware in a sandbox or virtual machine to monitor its actions during execution, such as system calls, file alterations, network interactions, and registry errors. Dynamic detection can detect malicious activities that are not picked up by static analysis by observing the activity of the malware during its execution.

2) Spam

In spam detection, SHAP explains why an email or message is classified as spam or legitimate by measuring the contribution of features such as suspicious keywords, URLs, sender information, and message characteristics. This helps identify the factors influencing predictions, improves model transparency, detects misclassifications, and increases the reliability of AI-based spam detection systems [30].

3) Botnet

The goal of a botnet assault is to compromise and degrade the victim's resources, such as 8 crashing websites, by continually executing malicious code on a network of interconnected machines. The following are the five stages that make up a typical botnet's lifetime: initial injection, secondary injection, connection, malicious activity, and maintenance and upgrading (Figure 3) [31].

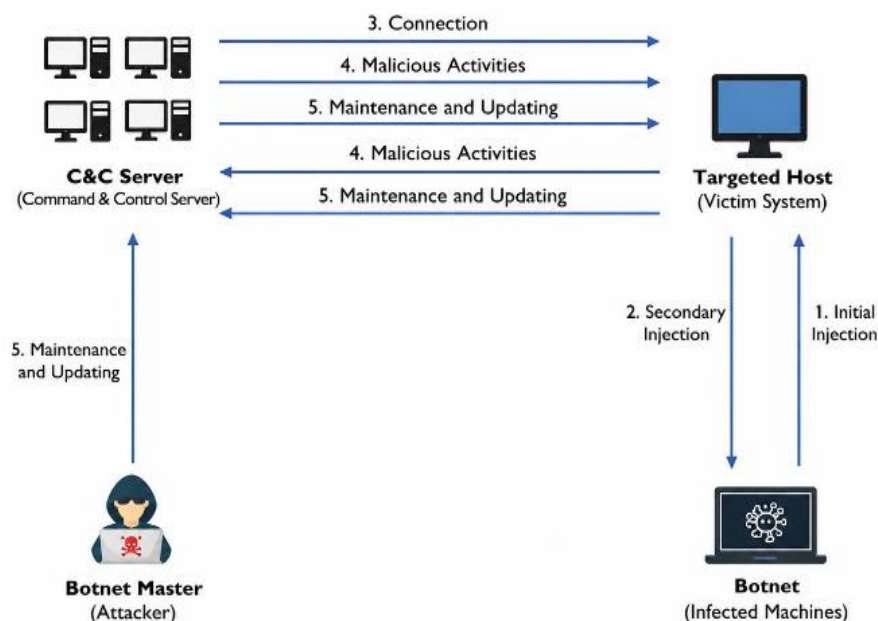


Fig. 3. The typical lifecycle of a botnet

4) Fraud

There was an upsurge in online financial fraud and hacking of personal accounts during the COVID-19 pandemic's most stringent shutdown periods [32]. As a result, AI systems targeted at fighting fraud may be a useful tool for many banks in this fight.

5) Phishing

The term "phishing" refers to the practice of sending phishing emails that seem like they came from a legitimate company [33]. Obtaining sensitive information, such as login credentials and financial data, or infecting the victim's machine with harmful malware is the objective. One sort of online fraud that is becoming more widespread is phishing websites. One useful method for detecting these sites is the multi-modal hierarchical attention model (MMHAM).

6) Network Intrusion

The term "network intrusion" describes an unlawful entry into a computer system inside an organization or a specific domain. Network IDS, on the other hand, are characterized as keeping an eye on system or network activity for signs of suspicious or harmful conduct that breaches security or industry standards. A lot of recent work has used ML and DL techniques to construct effective NIDSs [34]. Also, in an attempt to fortify NIDSs, a number of cybersecurity experts have dabbled with XAI, which grants black-box AI systems further explainability.

7) Domain Generation Algorithms (DGA)

A common tactic employed by DGAs, a kind of virus, is to generate an overwhelming amount of

domain names. These names can then be used to evade detection by Command and Control (C2) servers [35]. Traditional methods of domain protection, such as blacklisting or sink-holing, are ineffective against malicious domains because of the sheer number of available domain names. One typical tool for DGA dynamics is a seeded function. Identifying the virus, DGA, and seed value is the initial step for administrators in preventing a DGA approach. After that, they can exclude any networks or servers that may have been detrimental earlier in the chain.

8) Denial-of-Service (DOS)

As a big danger to the Internet, denial-of-service assaults have prompted several security proposals [36]. When malicious nodes repeatedly send out bogus messages in an effort to slow down or even completely disable other nodes' networks, this is known as a continuous denial-of-service assault. Traditional intrusion detection systems are finding it increasingly difficult to detect the more modern and complicated denial-of-service attacks due to the increased sophistication and diversity of these attacks over the last many years.

B. Challenges and Limitations of XAI in Cybersecurity

XAI in cybersecurity faces challenges related to complex data, model transparency, computational cost, privacy, and scalability. The models are complex and hard to interpret, and explanations could be subject to adversarial manipulation. The absence of uniform approaches to evaluation and real-time explanations also restricts trustworthy deployment in the dynamic cybersecurity domain.

1) XAI-Driven Security Threats

The information gained from XAI presents a fundamental challenge to the security of DL models, as an adversary can leverage this information in a manner beyond what was originally designed to be considered by the adversary, making their attack more sophisticated and adaptive. Explainability enables the adversary to know how and why a decision was made, allowing them to know the key elements for an attack and adapt accordingly with minimal effort [37]. For example, proved that human explanation of LIME was usable to infer decision boundaries for authentication systems, and an attacker could pose a query that was very similar to the legitimate user's one, and hence achieve high identification rates.

2) XAI-Driven Privacy Risks

XAI attacks go beyond compromising models, they also pose a threat to the privacy of the data, particularly in the context of model MEA and MIA. By incorporating transfer learning, XaMEA is able to achieve up to 13.6% extraction accuracy on average, outperforming baseline extraction methods and highlighting the nature of the privacy risks posed by transparent explanation mapping. Likewise, MEGEX is an exact gradient-based model extraction method that adds data-free explanations without requiring any real input samples, increasing fidelity [38]. These attacks show that an adversary's ability to reverse-engineer a model's decision limits is directly proportional to the completeness of the explanatory information.

3) Robustness of XAI Techniques

New attack surfaces can be introduced via XAI approaches, as mentioned in the preceding section. Attackers may trick an IDS, fool a model, and baffle analysts, according to recent study [39]. One might, for instance, manipulate the values of the inputs to make the attributions made by SHAP and LIME consistent with the data but incorrect. Analogously, interpretable models, such as DTs can be reverse-engineered for poisoning attacks.

4) Lack of Cybersecurity Evaluation Metrics

A key challenge is the need for a clear and concise domain-specific assessment criteria that can fully represent what a successful explanation is for cybersecurity practitioners [40]. In order to evaluate explanations for activities such as monitoring network traffic, reacting to intrusion detection warnings, and identifying phishing attempts, specific types of explainability are required. Normal metrics (such as precision) frequently fall short when it comes to meeting these demands.

4. Literature Review

The use of XAI in intelligent cybersecurity is on the rise, according to recent research. XAI improves decision-making, trust, and transparency by utilizing DL models, Grad-CAM, SHAP, and LIME.

A. Vashisth, et al. (2026) systematic of the threat scenario of cybersecurity in Intelligent Transportation Systems is provided with references to attack models, vulnerabilities, threat actors, and defenses. The structured taxonomy is created to classify risks at the perception, communication, network, and application layers. examines the current research works, juxtaposes the available methods of protection, and assesses their drawbacks. Also notes two new challenges that include real-time attack detection, preservation of privacy, adversarial threat of AI, and propagation of attack across domains [41].

R. H. M and S. J. Upadhyaya (2025) utilize SHAP, LIME, and post hoc explanation counterfactual approaches to assist people in understanding explanations with an optimistic outlook, and evaluate detection models—RF, GB, CNN, and BiLSTM—on publicly accessible datasets, namely CICIDS2017 and UNSW-NB15. BiLSTM achieved the highest detection accuracy score of 96.3%, while RF exhibited the best performance-to-explainability ratio. It is a real-time analyst dashboard that helps to display the model findings [42].

M. T. Masud et al. (2025) delve into the following XAI-related ideas. The latest findings on XAI are summarized in this section. The article delves into recent taxonomies that provide light on various ML techniques. Among them are DL methods that have been created and researched in different IoT classification systems. Experts need more complex results from AI models than just binary ones in order to enable the cyber resilience of IoT systems, making AI model outputs critical for cybersecurity. One way to summarize the challenges and shortcomings in XAI when it comes to cybersecurity is to look at the existing applications and safety-related threat models that attempt to explain resistance towards IoT devices [43].

S. Safavi, et al. (2025) The suggested paradigm combines SHAP, LIME, and Grad-CAM, among others, to demonstrate how AI decision-making is rationalized. Data protection, adversarial mitigation, and model provenance are few of the comprehensive security procedures it employs to guarantee the stability and dependability of AI systems. The use of trust-enhancing techniques including thorough logging, human-in-the-loop monitoring, and model certification further supports AI-driven cybersecurity judgments [44].

R. Barton, et al. (2024) an analysis of recent advances in artificial intelligence (AI) for cybersecurity purposes, specifically focusing on concerns regarding the

widespread adoption of AI. Kindly furnish research domains and pertinent, current publications covering the domain of AI in cybersecurity from January 2022 to January 2024. In order to educate AI about privacy, new AI datasets, such real-time and privacy datasets, are required, as shown by the comparison and contrast. The lack of clarity and knowledge around the reasoning behind algorithmic decisions and their results necessitates advancements in XAI [45].

K. Kumar, et al. (2023) explore these relationships and reveal that e-Governance partially plays a mediating role between AI and cybersecurity, where stakeholders play an influencing role. The growing use of AI in everyday life, from healthcare to finance to surveillance, highlights the need for data privacy. This poses new privacy and ethical issues when dealing with large amounts of data and requires robust security

measures. The abstract concludes by pointing out how GenAI models likely affect society, ethics, and privacy in 2023 through their impact on the cybersecurity scene [46].

Despite the remarkable progress in the field of AI-driven cybersecurity, the research is fragmented into three key areas: threat detection, explainable AI, IoT security, and intelligent transportation systems. The work investigating accuracy of detection is mostly not real-time, not preserving privacy, not adversarial and not trustworthy for decision-making. An all-encompassing and adaptive framework is essential for cybersecurity, however, it must be AI-powered. Recent studies revealed trends and insights, identifying critical challenges and future directions, as detailed in Table I, and highlighting the need for scalable, privacy-preserving, trustworthy, and real-time AI-driven cybersecurity solutions

TABLE I. EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR INTELLIGENT CYBERSECURITY APPLICATIONS

Reference	Study On	Approach	Key Findings	Challenges	Future Direction
A. Vashisth et al. (2026)	Cybersecurity threats in Intelligent Transportation Systems	Systematic review and structured threat taxonomy	Classified threats across perception, communication, network, and application layers	Real-time attack detection, privacy preservation, adversarial AI, and cross-domain attacks	Develop real-time, privacy-preserving, and adaptive cybersecurity frameworks
R. H. M. & S. J. Upadhyaya (2025)	Explainable AI for intrusion detection	RF, Gradient Boosting, CNN, BiLSTM, SHAP, LIME, and counterfactual explanations	BiLSTM achieved 96.3% accuracy, while RF provided better explainability	Limited real-time validation, scalability, and interpretability	Develop scalable and real-time XAI-based intrusion detection systems
M. T. Masud et al. (2025)	XAI for IoT cybersecurity	XAI taxonomies and AI-based cybersecurity methods	Emphasized explainability for cyber-resilience and expert decision-making	Scalability, reliability, and practical deployment challenges	Develop reliable and scalable XAI frameworks for IoT security
S. Safavi et al. (2025)	Trustworthy AI-driven cybersecurity	SHAP, LIME, Grad-CAM, data protection, adversarial mitigation, and human-in-the-loop	Improved explainability, security, model provenance, and trust	Limited real-world evaluation and adaptability to evolving threats	Develop adaptive, trustworthy, and robust AI cybersecurity solutions
R. Barton et al. (2024)	AI advancements in cybersecurity	Structured literature review of AI cybersecurity research	Identified the need for real-time and privacy-focused datasets and better XAI	Lack of real-time datasets, privacy datasets, and transparent AI decisions	Create high-quality datasets and improve explainable AI capabilities
K. Kumar et al. (2023)	AI, cybersecurity, privacy, and e-Governance	Analysis of AI-cybersecurity relationships	Highlighted privacy, ethical, and societal challenges of AI adoption	Data privacy, ethical concerns, and governance challenges	Establish stronger privacy safeguards, ethical guidelines, and secure AI governance

5. Conclusion and Future Work

AI is an essential technology that can improve the capabilities of modern cybersecurity solutions, offering intelligent threat detection, automated response, and proactive defense against new cyber threats. Despite their success, conventional AI models have also been criticized for their lack of interpretability, however, this has led to a demand for XAI to provide transparency

and trustworthy decision-making. The survey has been designed to provide a comprehensive overview of XAI for intelligent cybersecurity applications, ranging from fundamentals of cybersecurity and explainable AI to intelligent threat detection and response solutions, essential cybersecurity applications, and recent research efforts. The comparative analysis highlighted the potential of using XAI techniques such as SHAP, LIME, Grad-CAM, and counterfactual explanations to enhance

model transparency, increase analysts' trust in the models, and facilitate informed security decisions. The survey also showed that privacy concerns, adversarial attacks, robustness, scalability, and a lack of standardized evaluation metrics continue to be significant challenges that limit the practical application of XAI in real cybersecurity applications.

Future work should focus on developing a scalable, privacy-preserving and adversarial-robust XAI framework for intelligent Cybersecurity. Adopting real-time explainability, creating standardized assessment metrics, and piloting solutions with diverse real-world datasets will further bolster the explainability and trustworthiness of AI-driven cybersecurity solutions.

REFERENCES

1. M. Humayun, M. Niazi, N. Z. Jhanjhi, M. Alshayeb, and S. Mahmood, "Cyber Security Threats and Vulnerabilities: A Systematic Mapping Study," *Arab. J. Sci. Eng.*, vol. 45, no. 4, pp. 3171–3189, 2020, doi: 10.1007/s13369-019-04319-2.
2. P. Kumar, "A Zero Trust-Based Approach to Modern Cybersecurity Challenges in Software Development," *Int. J. Emerg. Res. Eng. Technol.*, vol. 6, no. 9, pp. 113–122, September, 2025.
3. S. Kumara and M. Shah, "Securing national connectivity infrastructure through identity resilience: implications for zero trust," *Futur. Technol. Open Access J.*, vol. 5, no. 3, pp. 276–286, August, 2026, doi: 10.55670/fppl.futech.5.3.24.
4. J. Kaur and K. R. Ramkumar, "The recent trends in cyber security: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5766–5781, Sep. 2022, doi: 10.1016/j.jksuci.2021.01.018.
5. B. Guembe, A. Azeta, S. Misra, V. C. Osamor, L. Fernandez-Sanz, and V. Pospelova, "The Emerging Threat of AI-driven Cyber Attacks: A Review," *Appl. Artif. Intell.*, vol. 36, no. 1, p. 2037254, Dec. 2022, doi: 10.1080/08839514.2022.2037254.
6. A. Kuppa and N.-A. Le-Khac, "Black Box Attacks on Explainable Artificial Intelligence(XAI) methods in Cyber Security," in 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9206780.
7. V. Rohilla, K. Rohilla, P. Kumar, and N. Jain, "Intrusion Detection in Network Traffic Using Feature Selection and Optuna-Based Machine Learning Optimization," in 2026 3rd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India: IEEE, Apr. 2026, pp. 1–6. doi: 10.1109/RMKMATE69073.2026.11518834.
8. S. C. Pallaprolu, "Zero-day Attack Identification in Streaming Data: Nearest Neighbor Heuristics and Dynamic Semantic Network Generation in the Spark Eco-system," University of Maryland, Baltimore County, 2017.
9. I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: an overview from machine learning perspective," *J. Big Data*, vol. 7, pp. 1–29, 2020.
10. S. Wali and I. Khan, "Explainable AI and Random Forest-Based Reliable Intrusion Detection System," *TechRxiv*, vol. 2021, no. 1218, 2021, doi: 10.36227/techrxiv.17169080.v1.
11. S. Singh, "Advancing Network Security in 5G: Leveraging the 5G-NIDD Dataset for Intrusion Detection and Mitigation," in 2025 IEEE 12th International Conference on Cyber Security and Cloud Computing (CSCloud), New York City, NY, USA: IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/CSCloud66326.2025.00055.
12. S. P. Sivashanmugam, S. Kumara, A. Mohile, and D. R. Suram, "Improving Detection Accuracy of Phishing Attacks with Feature Selection and Machine Learning Techniques in Cybersecurity," in 2026 1st International Conference on Emerging Trends in Advancements and Applications of Computational Intelligence Techniques (ETAAC), Bhubaneswar, India: IEEE, 2026, pp. 1–6, April. doi: 10.1109/ETAAC69135.2026.11542138.
13. V. Jain and A. Mitra, "Real-Time Threat Detection in Cybersecurity," in Machine Intelligence Applications in Cyber- Risk Management, 2024, pp. 315–344. doi: 10.4018/979-8-3693-7540-2.ch014.
14. S. C. Pallaprolu, R. Sankineni, M. Thevar, G. Karabatis, and J. Wang, "Zero-Day Attack Identification in Streaming Data Using Semantics and Spark," in 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, HI, USA: IEEE, Jun. 2017, pp. 121–128. doi: 10.1109/BigDataCongress.2017.25.
15. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, Jul. 2009, doi: 10.1145/1541880.1541882.
16. J. Sehgal, "Enhancing Site Reliability Engineering: Scalable Strategies for Automated Incident Response and System Resilience," *J. Artif. Intell. Mach. Learn. Data Sci.*, vol. 2, no. 4, pp. 2484–2488, 2024, doi: 10.51219/jaimld/jaya-sehgal/533.
17. V. Sai Swaroop Reddy, "Cybersecurity Threat

- Prediction Using Machine Learning," *Int. J. Sci. Res.*, vol. 12, no. 4, pp. 1972–1976, 2023, doi: 10.21275/sr23048115831.
18. S. Cramer, C. Glantz, G. Landine, S. Eggers, and J. Knight, "Technical Guide for Implementing Cybersecurity Continuous Monitoring in the Nuclear Industry," 2021.
19. V. Koppireddy, D. G. Parasad, D. Patel, S. R. Somula, V. Kamaraj, and A. Meher, "AI Security Governance for Trustworthy Intelligent Systems: A Critical Analysis of Cyber Risks, Privacy Threats, and Responsible Deployment Frameworks," *Int. J. Res. Trends Innov.*, vol. 11, no. 07, pp. a407–a435, July, 2026, doi: 10.56975/ijrti.v11i7.214029.
20. V. K. Sharma and K. S. Abhilash, "Latency-Aware Edge-Cloud Architecture for 5G IoT Integration," in 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India: IEEE, Sep. 2025, pp. 1398–1405. doi: 10.1109/ICESC65114.2025.11212232.
21. M. Özkan-Okay, R. Samet, Ö. Aslan, and D. Gupta, "A Comprehensive Systematic Literature Review on Intrusion Detection Systems," *IEEE Access*, vol. 9, pp. 157727–157760, 2021, doi: 10.1109/ACCESS.2021.3129336.
22. V. B. Savant and R. D. Kasar, "A Review on Network Security and Cryptography," *Res. J. Eng. Technol.*, vol. 3, no. 1, pp. 110–114, 2021, doi: 10.52711/2321-581x.2021.00019.
23. S. Singh, Y.-S. Jeong, and J. H. Park, "A survey on cloud computing security: Issues, threats, and solutions," *J. Netw. Comput. Appl.*, vol. 75, pp. 200–222, Nov. 2016, doi: 10.1016/j.jnca.2016.09.002.
24. M. Moucharraf, M. Ridouani, F. Salahdine, and N. Kaabouch, "IoT Security: A Comprehensive Review of Architectures, Threat Models, Detection Methods, and Countermeasures," *Futur. Internet*, vol. 18, no. 5, pp. 1–31, 2026, doi: 10.3390/fi18050266.
25. V. K. R. K. Koppireddy, "Passwordless Authentication: Enhancing Security And User Experience In Modern Digital Ecosystems," *Int. J. Inf. Technol. Manag. Inf. Syst.*, vol. 16, no. 1, pp. 82–96, Feb. 2025, doi: 10.34218/IJITMIS_16_01_008.
26. J. Glöckler, J. Sedlmeir, M. Frank, and G. Fridgen, "A Systematic Review of Identity and Access Management Requirements in Enterprises and Potential Contributions of Self-Sovereign Identity," *Bus. Inf. Syst. Eng.*, vol. 66, no. 4, pp. 421–440, 2024, doi: 10.1007/s12599-023-00830-x.
27. D. Vale, A. El-Sharif, and M. Ali, "Explainable artificial intelligence (XAI) post-hoc explainability methods: risks and limitations in non-discrimination law," *AI Ethics*, vol. 2, no. 4, pp. 815–826, 2022, doi: 10.1007/s43681-022-00142-y.
28. L. Maglaras, V. C. Gerogiannis, M. A. Ferrag, A. Lekidis, A. Lakas, and N. Moradpoor, "The Dual Role of Artificial Intelligence and LLM in Cybersecurity," in 2025 International Conference on Artificial Intelligence Security and Governance (ICAISG), IEEE, Dec. 2025, pp. 113–118. doi: 10.1109/ICAISG68699.2025.11452144.
29. T. K. Lengyel, S. Maresca, B. D. Payne, G. D. Webster, S. Vogl, and A. Kiayias, "Scalability, fidelity and stealth in the DRAKVUF dynamic malware analysis system," in Proceedings of the 30th Annual Computer Security Applications Conference, New York, NY, USA: ACM, Dec. 2014, pp. 386–395. doi: 10.1145/2664243.2664252.
30. E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems," *Heliyon*, vol. 5, no. 6, Jun. 2019, doi: 10.1016/j.heliyon.2019.e01802.
31. S. S. C. Silva, R. M. P. Silva, R. C. G. Pinto, and R. M. Salles, "Botnets: A survey," *Comput. Networks*, vol. 57, no. 2, pp. 378–403, Feb. 2013, doi: 10.1016/j.comnet.2012.07.021.
32. D. Buil-Gil, F. Miró-Llinares, A. Moneva, S. Kemp, and N. Díaz-Castaño, "Cybercrime and shifts in opportunities during COVID-19: a preliminary analysis in the UK," *Eur. Soc.*, vol. 23, no. S1, pp. S47–S59, 2021, doi: 10.1080/14616696.2020.1804973.
33. A. Das, S. Baki, A. El Aassal, R. Verma, and A. Dunbar, "SoK: A Comprehensive Reexamination of Phishing Research From the Security Perspective," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 1, pp. 671–708, 2020, doi: 10.1109/COMST.2019.2957750.
34. G. Andresini, A. Appice, F. P. Caforio, D. Malerba, and G. Vessio, "ROULETTE: A neural attention multi-output model for explainable Network Intrusion Detection," *Expert Syst. Appl.*, vol. 201, p. 117144, Sep. 2022, doi: 10.1016/j.eswa.2022.117144.
35. P. Wagner et al., "PTB-XL, a large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, no. 1, p. 154, 2020, doi: 10.1038/s41597-020-0495-6.
36. R. H. Jhaveri, S. J. Patel, and D. C. Jinwala, "DoS Attacks in Mobile Ad Hoc Networks: A Survey," in 2012 Second International Conference on Advanced Computing & Communication Technologies, IEEE, Jan. 2012, pp. 535–541. doi: 10.1109/ACCT.2012.48.

37. W. Garcia, J. I. Choi, S. K. Adari, S. Jha, and K. R. B. Butler, "Explainable Black-Box Attacks Against Model-based Authentication," pp. 1–23, 2018. doi: 10.1109/IEIR62538.2024.10959777.
38. A. Yan, T. Huang, L. Ke, X. Liu, Q. Chen, and C. Dong, "Explanation leaks: Explanation-guided model extraction attacks," *Inf. Sci. (Ny).*, vol. 632, pp. 269–284, Jun. 2023, doi: 10.1016/j.ins.2023.03.020.
39. J. Cohen, E. Rosenfeld, and J. Z. Kolter, "Certified adversarial robustness via randomized smoothing," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 2323–2356, 2019.
40. D. Bhusal et al., "SoK: Modeling Explainability in Security Analytics for Interpretability, Trustworthiness, and Usability," in *Proceedings of the 18th International Conference on Availability, Reliability and Security*, in *ARES '23*. New York, NY, USA, NY, USA: ACM, Aug. 2023, pp. 1–12. doi: 10.1145/3600160.3600193.
41. A. Vashisth, G. Kaur, A. Sharma, B. Kaur, and M. Rakhra, "Cybersecurity Threat Landscape in Intelligent Transportation Systems: A Systematic Review," in *2026 International Conference on Intelligent Systems in Engineering, Secured Systems and Cybersecurity (ICISESSC)*, IEEE, Apr. 2026, pp. 770–775. doi: 10.1109/ICISESSC68634.2026.11542737.
42. R. H. M and S. J. Upadhyaya, "Explainable AI (XAI) for Cyber Threat Intelligence and Decision Making," in *2025 International Conference on Transformative Computing Technologies (ICTCT)*, IEEE, Sep. 2025, pp. 79–84. doi: 10.1109/ICTCT69201.2025.00030.
43. M. T. Masud, M. Keshk, N. Moustafa, I. Linkov, and D. K. Emge, "Explainable Artificial Intelligence for Resilient Security Applications in the Internet of Things," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 2877–2906, 2025, doi: 10.1109/OJCOMS.2024.3413790.
44. S. Safavi, M. S. H. Abdulnabi, M. E. Rana, and S. Alizadeh, "From Black Box to Trustworthy AI: A Secure Framework for Explainable Cybersecurity Decision-Making," in *2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, IEEE, May 2025, pp. 1–4. doi: 10.1109/ASSIC64892.2025.11158699.
45. R. Barton, P. W. C. Prasad, I. Seher, and A. Elchouemi, "Artificial Intelligence (AI) in Cybersecurity and Inhibitors to AI Adoption," in *2024 International Conference on Intelligent Education and Intelligent Research (IEIR)*, IEEE, Nov. 2024, pp. 1–10. doi: 10.1109/IEIR62538.2024.10959777.
46. K. Kumar, Kuldeep, and B. Bhushan, "Augmenting Cybersecurity and Fraud Detection Using Artificial Intelligence Advancements," in *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, IEEE, Nov. 2023, pp. 1207–1212. doi: 10.1109/ICCCIS60361.2023.10425069.