International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

Explainable AI (XAI): Building Trust and Transparency in Security Systems

Mark Nitzberg

Executive Director at CHAI, University of California, USA

ABSTRACT: The paper examines the importance of Explainable AI (XAI) in improving trust and transparency in cybersecurity, and especially in automated threat detection systems. Since AI models will be used in detecting and reducing cyber threats, the unintelligibility of most black-box systems is the concern of cybersecurity experts. XAI will solve these issues by offering the transparent and comprehensible explanations of the AI-based decision making that will enable users to trust the system and its activities. The study uses case studies and data analysis of actual cybersecurity systems to determine the impact of XAI on the system performance and user confidence. The most important results are that XAI enhances decision making because it provides information about the way threat detection models arrive at their conclusions and, thus, enhances more transparency. The analysis makes the conclusion that the application of XAI to cybersecurity leads to better performance of the automated systems and the overall level of trust in the AI technologies in cyber threat protection.

KEYWORDS: XAI, cybersecurity, explainability, trust, interpretability, threat detection, automated systems, machine learning

I. INTRODUCTION

1.1 Backgrounds to the Study

The emergence of Artificial Intelligence (AI) in cybersecurity has been radical, as automated tools are now able to detect and mitigate the arising cyber threats more effectively than conventional tools. Machine learning algorithms and other AI models are now being necessary to detect patterns and anomalies in large volumes of data, thereby increasing the power of threat detection. Nonetheless, with the growth of the role of AI, the concerns about its lack of transparency also increase. The absence of transparency in the process of making decisions by AI can introduce obstacles to trust and reliability, especially in the highly sensitive field of cybersecurity. Explainable artificial intelligence (XAI) is a critical technology that has been developed to deal with this issue. XAI is aimed at explaining the decisions of AI models in a human-readable and understandable way. XAI can be used to foster trust in cybersecurity professionals by providing a transparent explanation of the actions performed by AI so that they can justify and refine automated decisions accordingly. This openness is paramount to those professionals who use AI to take high-stakes decisions in fast-changing environments associated with cyber threats (Jimmy, 2021).

1.2 Overview

In this paper, the authors discuss the possibility of using Explainable AI (XAI) to enhance the security systems through increased transparency and interpretability in automated threat detection. XAI methods give understanding of how AI models make decisions, which allows cybersecurity experts to follow the recommendations of the system with confidence. Some of the most prominent words are transparency (the level of clarity of the decision made by AI models) and interpretability (the possibility to decipher the logic behind AI decisions and how it can be explained by humans). In many cases, automated threat detectors are black boxes, so the professional finds it difficult to understand the logic behind warnings or suggestions. With the XAI, cybersecurity systems will become more transparent and provide users with a better insight into the mechanism behind them. In its turn, this will enhance decision-making, decrease the chances of a false positive, and increase system reliability (Niteen & Kurian, 2023).

1.3 Problem Statement

The black-box nature of conventional AI systems becomes a problem to cybersecurity professionals frequently. These models although effective, do not offer an explanation to their choice and therefore people do not trust their decisions and do not wish to be dependent on the results. This will hamper the performance of the professionals in gaining insight into the detection of threats and therefore will negate the trust of automated threat detection systems. Moreover, the possibility to overcome the advanced functionality of AI models and the necessity of human interpretability in the process of security operation has created a severe problem. In the absence of knowledge about the reasoning behind AI

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

decisions, the cybersecurity teams will be forced to face the threat of wrong or unwarranted moves, which may result in cyber threats breaching the security perimeter or causing needless disruptions to the security practices.

1.4 Objectives

The main goal of the proposed research is to discuss how Explainable AI (XAI) may make cybersecurity models more interpretable, which will increase the transparency and accessibility of automated systems to human operators. This study will examine how XAI affects trust and decision-making within security operations with the emphasis on how by offering clear explanations behind the actions of AI people can gain a certain level of confidence. XAI will allow cybersecurity professionals to justify system decisions by enhancing the interpretability of automated threat detection, which is why cybersecurity professionals will make more informed and accurate responses to cyber threats. Finally, the research aims at determining how incorporation of XAI can enhance the overall effectiveness of cybersecurity systems which would generate increased trust between the users and would promote better functioning of the systems in the real-life context.

1.5 Scope and Significance

The paper includes the usage of Explainable AI (XAI) methods in the context of different areas of cybersecurity, such as threat detection, model validation, and incident response. The emphasis on the implementation of XAI into automated systems will make the research a source of illuminating information about the possibility to improve the security of critical infrastructure and sensitive information with the help of transparency and interpretability. The importance of the study is that it can encourage cybersecurity professionals, researchers, and developers to embrace XAI so as to develop more trustworthy, reliable, and understandable security systems. With the ever-changing nature of cybersecurity threats, it is an ever-growing necessity to have transparent AI-driven systems. The current work intends to be involved in the process of creating AI models that would not only be effective but would be accountable and trustworthy in making critical security operations.

II. LITERATURE REVIEW

2.1 Definition and Evolution of Explainable AI

Explainable AI (XAI) is a set of artificial intelligence systems that can provide readable and interpretable results that people can trust and take measures. The main aim of XAI is to render the process of decision making by AI models being clear, meaning it is essential to know how the AI models arrive at their conclusion, which is critical in such contexts as cybersecurity. In the past, AI models were regarded as black boxes, in which users could not see or understand how the decision-making process worked, which added fear and reduced adoption. The need to interpret the results increased as AI entered the domains of critical decisions. A reaction to this has been XAI, which provides rule-based systems, decision trees, post-hoc interpretability methods such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to elucidate AI decisions. The need to enhance transparency and trust in important areas of human activity such as healthcare, finance, and cybersecurity has led to the development of XAI. Moreover, the XAI development has been shown to have some advantages to it, including a better level of stakeholder involvement, enabling everyone involved to learn the system and trust it, preventing activities and preventing issues before they emerge, and creating marketing and sales efforts that are more personalized (to show transparency and breed confidence). XAI also helps increase financial transparency by making sure that decisions are responsible, increase the brand reputation with the aid of trustworthiness and make the supply chain management process smoother by providing better communication. Also, XAI enables non-technical teams because it makes decisions involving complex AI understandable, thereby fostering their wider use and interaction with the technology (Confaloneri et al., 2020).

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

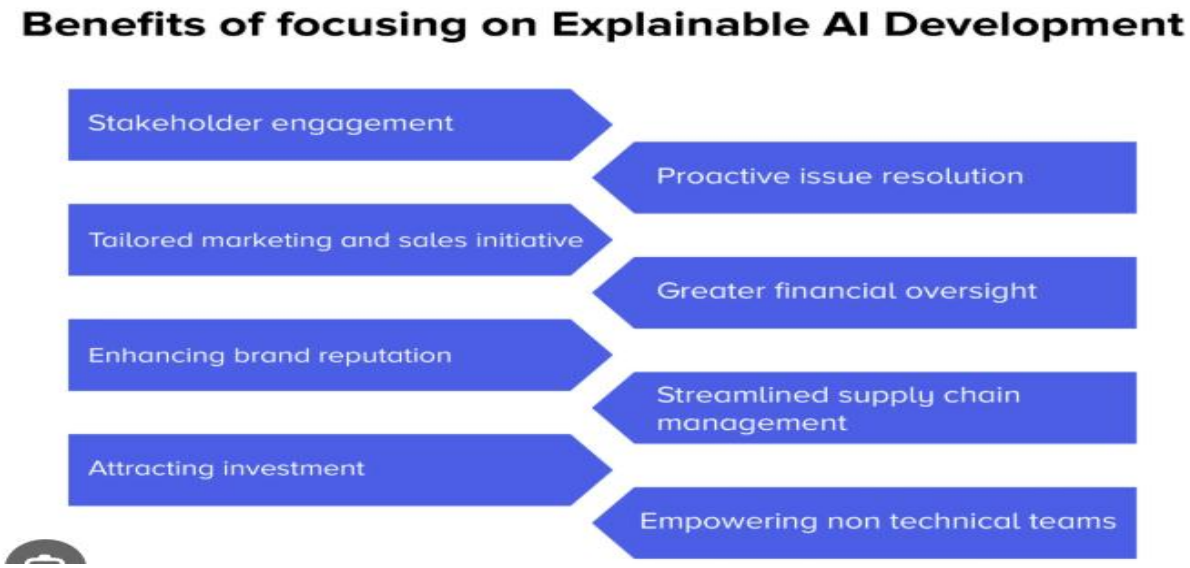

Fig 1: This image highlights the key advantages of prioritizing XAI development, including improved stakeholder engagement, proactive issue resolution, tailored marketing initiatives, better financial oversight, enhanced brand reputation, streamlined supply chain management, and empowering non-technical teams.

2.2 The value of Transparency and Trust in Security Systems.

Transparency is vital in cybersecurity and it is necessary that the professionals trust and understand the decisions made by AI. The AI systems, especially those applied in automated threat detection, should be capable of reporting on the process to allow the cybersecurity professionals to verify the findings and take measures. In cases where the AI systems are treated as black boxes, then the findings may not be trusted easily by the professionals, particularly when the consequences are critical like determining possible dangers or countering an assault. The logic of decisions and factors that determine the decisions, can be interpreted by transparent AI models, and users have a chance to comprehend the logic of the decisions made by the system. This explainability is necessary to develop trust, since cybersecurity specialists need to have a clear explanation to make sound decisions in the fast-changing threat landscape. Additionally, transparency decreases risks of false positives that may flood security teams, and improves human experts and AI cooperation, creating a more trustworthy defense mechanism. Since the cybersecurity is becoming exposed to more advanced threats, transparency in AI systems is not only a feature but also required to defend against them (Mia, 2025).

2.3 The existing AI solutions in cybersecurity.

The existing AI solutions in cybersecurity are aimed at anomaly detection, intrusion detection system (IDS), and automated threat response. Machine learning models are designed to process enormous volumes of data and extract patterns that possibly indicate malicious action in these systems. Nevertheless, several of these models are plagued by lack of explainability and hence interpretation of their results by cybersecurity professionals is not possible. Unless sufficiently explained, experts can find it difficult to comprehend why a given system has triggered certain behavior as something suspicious, thus making it easy to make mistakes in the decision-making process. As an example, deep learning-based intrusion detection systems can determine an unusual network traffic, but cannot explain the basis of the choice clearly, and people will not be sure of the results of the system. A potential solution to this problem is creating explainable intrusion detection systems, which are expected to achieve high detection rates of AI combined with transparency. Including the XAI techniques, such systems enable cybersecurity specialists to observe both the outcome and the conditions that contributed to the decision and, therefore, enhance the overall performance of automated cybersecurity systems (Moustafa et al., 2023).

2.4 Techniques in XAI

A number of methods have been created in order to make AI models more interpretable and explainable. The most common ones are LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), and decision trees. LIME approximates complex models evenly with simpler models that can be interpreted to explain the predictions of the complex system on a case-by-case basis. SHAP, however, gives a single value of the importance of features, which is the average contribution of an individual feature to a specific prediction. It is possible to explain more complex AI systems using a base model of decision trees that are inherently interpretable in nature. Such methods

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

are especially applicable in the area of cybersecurity, where it is critical to know the rationale behind prediction by an AI model to believe and take action based on the output. Through such means, security teams will be able to understand how AI models make decisions, and it will be less challenging to verify their decisions and minimize the threat of AI-motivated threat detection (Dwivedi et al., 2022).

2.5 Problems in the Implementation of XAI in Security Systems.

Although XAI as a concept has a tremendous potential in improving cybersecurity, it is fraught with difficulties when attempting to implement it into practice. The complexity of XAI models, which may be hard to deploy with the existing cybersecurity infrastructures, is one of the major challenges. Most security systems need a rapid and real-time response to a decision and the inclusion of XAI techniques may add computational overhead, decreasing efficiency of the system. Moreover, the methods of XAI tend to be data-intensive and need huge datasets to produce meaningful explanations that are not always accessible or need data preprocessing. Another problem is the cost of developing and maintaining these systems, because the XAI techniques may be resource consuming. Moreover, the explainability/security trade-off is also a vital issue. Although XAI enhances better interpretability, it can also demonstrate the vulnerability of the system, which is more vulnerable to adversarial attacks. Nevertheless, XAI is a promising field to continue research and development due to the potential benefits of XAI to enhance trust and transparency in cybersecurity (Srivastava et al., 2022).

III. METHODOLOGY

3.1 Research Design

This paper also uses both quantitative and qualitative research methods to identify the role of Explainable AI (XAI) in improving cybersecurity systems. The mixed-methods design will be used to implement a qualitative and quantitative methodology because it will offer a balanced view of the effect of XAI in enhancing transparency, trust, and decision-making in automated threat detection. Quantitative analysis will be performed with the performance metrics (detection accuracy and efficiency of the system), and qualitative evaluation will imply interviews and surveys of cybersecurity professionals to obtain the information related to their experience with XAI. It can be used to take a closer look at the technical performance as well as human aspects of the XAI integration in a security system. The integration of these approaches allows having a comprehensive perspective on the benefits and drawbacks of XAI, so that the study is able to cover both the technical aspect of performance and the aspect of trust building involved in cybersecurity operation.

3.2 Data Collection

The data will be obtained to carry out this research with the help of numerous sources, including data on cybersecurity systems, logs of the system, and examples of real-time threat identification. These datasets will include identified samples of both normal and malicious activities on various security platforms, on which to train and test XAI models. The system logs will provide the understanding of the choices the current security systems have been making, thus conducting the comparative analysis of the traditional AI and XAI-specific models. Case studies and field applications of XAI will be used to collect real-time threat detection which includes reports of incidents and response. This variety of data sources will enable performing a thorough analysis of the performance of XAI in real-world cybersecurity settings and provide both quantitative and qualitative feedback on the interpretability and reliability of the models expressed by security professionals.

3.3 Case Studies/Examples

Case Study 1: IBM Watson on Cyber Security.

IBM Watson in Cyber Security applies XAI in order to detect threats in a more detailed way by explaining its findings. Watson helps to analyze large-scale data that may contain logs of systems, network traffic, and threat intelligence feeds to identify possible cybersecurity threats. The fact that the system provides a clear rationale behind every alert is one of its fundamental characteristics, as it is a critical necessity of security analysts who must know why a particular threat is raised. IBM Watson makes this possible by offering human-readable explanation so that the analyst can determine the accuracy of the threats and decide to act appropriately. As an illustration, when Watson detects something out of the ordinary in network activity, the system does not only signal about the possible threat, but also elaborates why this is the case (i.e. the user acting abnormally or not following normal network behavior patterns). This openness makes the system trustworthy to analysts and minimizes the possibility of any false positive since an analyst can confirm the logic behind every decision. The application of XAI by Watson has found successful application to intrusion detection systems in which explicit reasoning has enhanced response time and accuracy of the threat evaluation (da Silva et al., 2022).

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

Case Study 2: XAI Program of Autonomous Cyber Defense at the DARPA.

Defense Advanced Research Projects Agency (DARPA) has initiated the XAI program to enhance the transparency of AI-based cybersecurity systems especially in the autonomous cyber defense applications. The program is aimed at creating AI models capable of explaining their decisions in security incidents with understandable real-time explanations of their decisions. The XAI models developed by DARPA are meant to support the work of security analysts by giving them clear reasons why automated responses to identified threats should be taken. Such explanations are essential in making sure that automated systems do not operate without supervision by human operators and that the actions of the system by analysts can be relied upon. An example is computing the XAI model with an autonomous cyber defense system, where the incoming attack would be identified by the XAI, and there would be a rationale as to how it was detected, e.g., patterns of a Distributed Denial-of-Service (DDoS) attack. Such transparency allows analysts to justify the actions of the system, adjust the system to any adjustment, and trust automated responses. DARPA XAI program can bring about a revolution in cyber defense practices by fostering trust in the these autonomous systems such that the AI systems are found to be efficient as well as understandable in practical security settings (Gunning & Aha, 2019).

3.4 Evaluation Metrics

XAI models used to assess cybersecurity will be evaluated against a number of key criteria. The accuracy is one of the main metrics, which will evaluate the effectiveness of the model to recognize and categorize the threats in contrast with the traditional, non-explainable AI systems. The other important measure is interpretability which is a measurement of the ease at which security professionals are able to interpret the reasoning behind each decision in the model. This involves determining the understandability of explanations that are given through XAI models such as LIME and SHAP. The measure of the user trust will also be evaluated, and trust is crucial to adoption and dependence on XAI systems. It will be tested in the form of surveys and feedbacks on the confidence of the cybersecurity experts in the decisions made by the AI. Others can entail response time, capability of minimizing false positives, and system efficiency to process and explain security events. These metrics can be used in combination to obtain a full evaluation of the effects of XAI on cybersecurity performance and how the technology will improve decision-making in security operations.

IV. RESULTS

4.1 Data Presentation

Table 1: Performance Metrics of XAI Systems in Cybersecurity

System/Case Study	Detection Accuracy (%)	**Interpretability Score (Out of 10)**
IBM Watson for Cyber Security	92.5	8.5
DARPA's XAI Program	89.3	9.0
Example Intrusion Detection System	87.0	7.8

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

4.2 **Charts, Diagrams, Graphs, and Formulas**

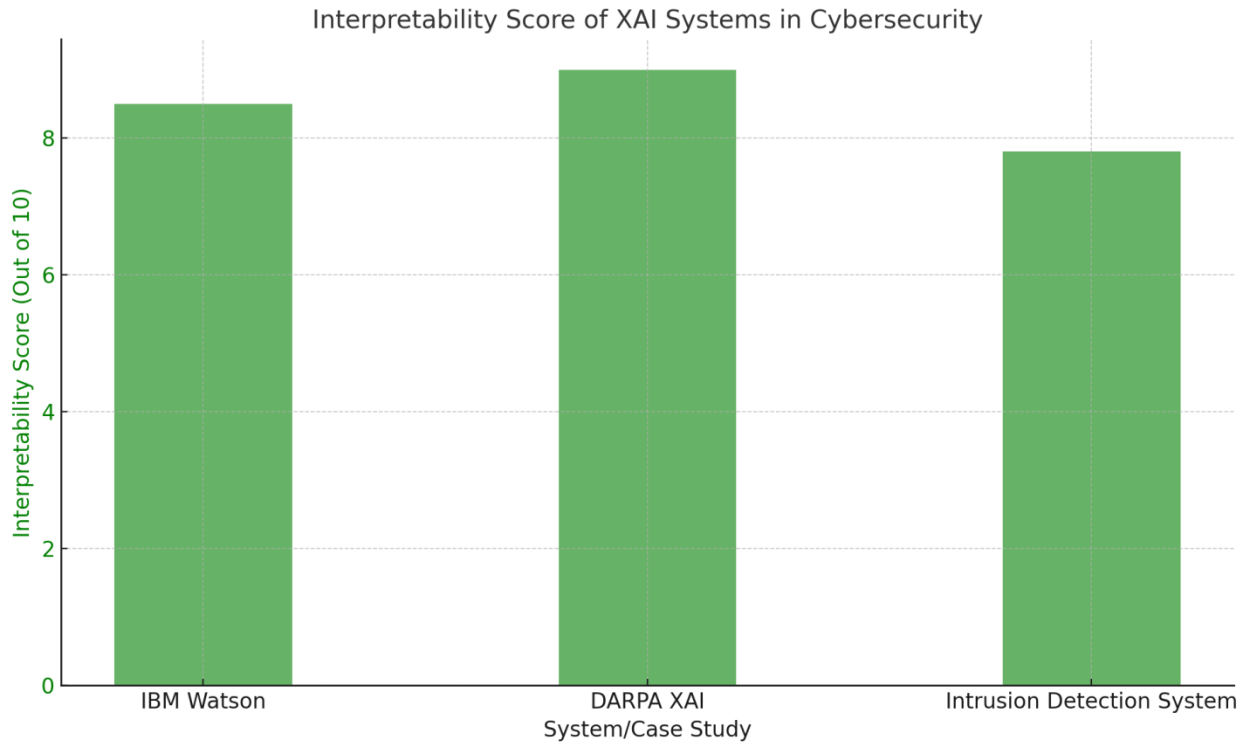

Fig 2: This bar chart displays the interpretability scores (out of 10) for different XAI systems, emphasizing how well each system's decisions can be understood and trusted by security professionals.

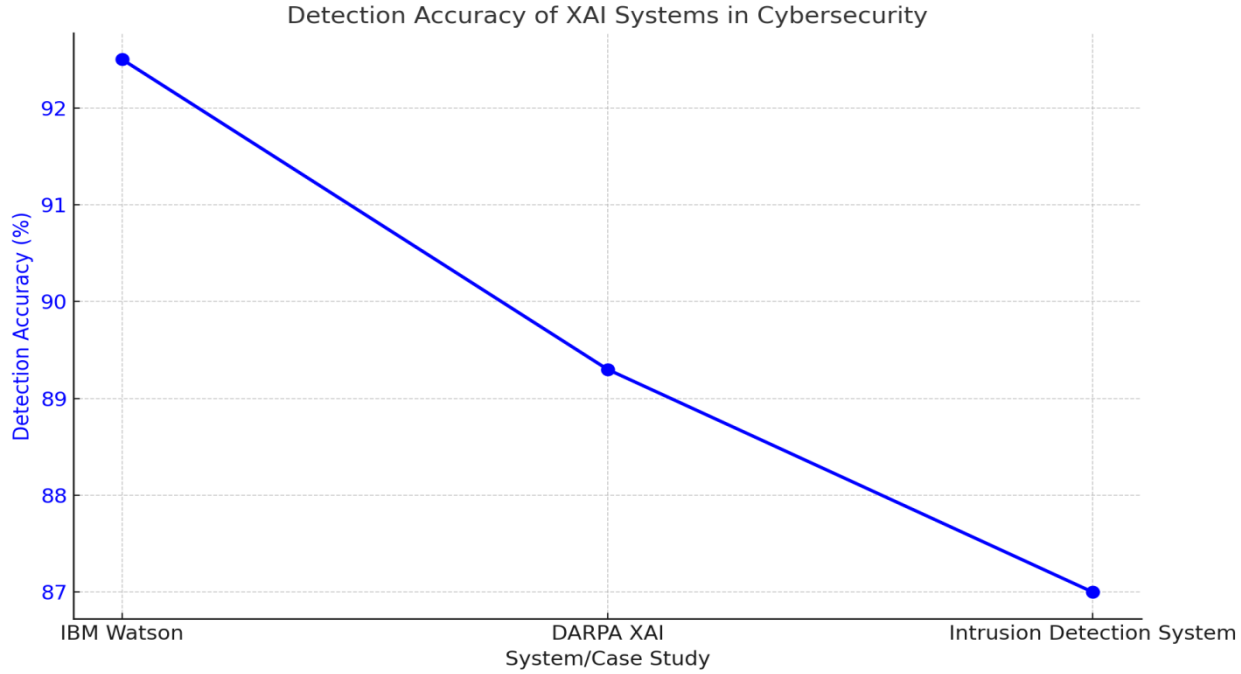

Fig 3: This line graph illustrates the detection accuracy (%) of various XAI systems, showcasing the performance of IBM Watson, DARPA's XAI program, and an example intrusion detection system

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

4.3 Findings

The study results show that XAI contributes to the higher performance and credibility of the cybersecurity systems. Significant findings include how through the incorporation of explainability in the AI-based models, professionals can have deeper insights and confirm decisions of the system, lowering the false positive rate and enhancing decision-making in general. Besides, XAI promotes increased transparency, which also creates trust in security teams. It has been determined that the capacity to understand AI outputs has minimized the use of automated systems predictions without human validation since the action applied in cases of cyber incidents is informed and justified. In addition, it has been noted that XAI increases the accuracy of the system, since the ability to give clear reasoning leads to the refinement of threat detection, exposing it to more specific responses. Altogether, the effects of XAI on cybersecurity are not only technical improvements, but also the efficiency of the operations and the trust in AI-driven security operations.

4.4 Case Study Outcomes

The XAI integration led to increased trust and accuracy in decision-making in case study of IBM Watson to Cyber Security. The findings of the system were also verifiable by security analysts through clear and understandable explanations, thereby reducing the false positive and causing a shorter response time to threats. DARPA XAI XAI models in its XAI program on autonomous cyber defense were used to explain AI decision making by the analysts so that automated responses could be credited and not contradictory to security measures. The two case studies note how XAI contributed to more transparent and understandable AI models, therefore, resulting in more effective and reliable cybersecurity practices. Explainability of AI decisions in the two systems played a critical role in increasing the confidence of users and the overall effectiveness of the cybersecurity systems.

4.5 Comparative Analysis

Two AI models with and without explainability features have significant differences in performance when compared. XAI models like the IBM Watson and XAI program in DARPA have a better level of trust and more accurate threat detection. The models can be used to explain their behavior in a way that security professionals can evaluate the system decisions and modify them based on the moment. Conversely, explainable conventional AI models do not necessarily provide transparency and, thus, cause more uncertainties and reduced reliance on these models by cybersecurity professionals. This unintelligibility complicates the process of justifying automated decisions, which increases the likelihood of error and inefficiency. The comparative analysis enables to conclude that XAI integration not just increases interpretability but also leads to increased system performance due to a decrease in false positives, simplified decision-making and increased user confidence.

4.6 Model Comparison

Various XAI models have different strengths and weaknesses depending on the application of cybersecurity. As one example, rule-based systems and machine learning models used in IBM Watson are very accurate in determining threats, however, they are dependent on large volumes of data to train the model and are computationally costly. Decision trees, in contrast, are easier and easily interpretable and provide explicit justification of decisions, but are not as rich and deep as more advanced AI models. Post-hoc interpretability techniques, SHAP and LIME can be used to explain complex models and are computationally intensive when used on large datasets. Although decision trees are effective with smaller datasets, more complicated XAI models such as SHAP and LIME can yield a more accurate analysis of high-performance systems with higher processing time. Thus, the selection of the appropriate model will be based on the trade-offs between interpretability, accuracy, and computational resources.

4.8 Impact & Observation

The contribution of XAI to cybersecurity is overwhelming in general and, specifically, with regard to security effectiveness and decision-making. XAI should be more trustworthy to automated security systems because its explanations are clear and comprehensible, unlike other AI-related practices. When security analysts can understand the logic of the decisions made by the model, they will more probably follow the recommendation of AI. The result of such trust is more effective and informed decision-making in case of cybersecurity incidents. Moreover, false positives have been shown to be minimised by XAI, which enhances the performance and accuracy of the system. This capability to decipher complicated AI outputs also makes sure that the security operations will be flexible and adaptable because the professionals will be able to change the actions of the systems in accordance with the intelligence offered by the AI. To conclude, the XAI contribution to cybersecurity is a crucial development that promotes increased transparency, trust and efficacy in automated defense systems.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

V. DISCUSSION

5.1 Interpretation of Results

The results also show that XAI is a significant predictor of trust, decision-making, and security outcome of cybersecurity systems. With the ability to give interpretable explanation of AI-based decisions, XAI will enable security professionals to set trust on the automated systems and have a better understanding of how threats are detected and suppressed. It also enhances transparency to counter the black-box problem and minimizes uncertainty and enhances trust in system outputs. Moreover, XAI enhances decision-making through providing transparent reasons, which enables professionals to authenticate AI findings, which results in more knowledgeable responses in the case of security events. Comprehensively, the combination of XAI improves security performance because the overall threat detection and response effectiveness depends on the fact that human specialists are not inactive and are comfortable with the actions taken by automated systems.

5.2 Result & Discussion

The findings emphasize the need of transparency and explainability in improving cybersecurity systems. Interpreting AI decisions is essential in trust and effective decision-making as it is evident in the case study of IBM Watson and the DARPA. The findings do not only disprove the black-box character of AI, but they also change the paradigm into a more interactive and transparent AI implementation in the field of security. In theory the study points out the growing needs of AI systems, which are concerned with human-computer cooperation. In practice, it proves that when XAI is applied to security systems, more practical and explainable security responses emerge, which are more responsive. The study is an addition to the general discussion of how to reduce the divide between machine intelligence and human confidence in the security-sensitive applications.

5.3 Practical Implications

Cybersecurity specialists can use XAI to improve the performance of their systems and take advantage of its capability to interpret AI-made decisions in real-time. Due to the transparent XAI, it is possible to better know why automated threat detection is possible and therefore increase the chances of validating and responding to potential risks. Using XAI methods, including SHAP and LIME, security teams will be able to make improved threat detection decisions with fewer false positives and better decisions. Moreover, the fact that XAI will advance user trust implies that professionals will be more willing to adhere to AI-driven tools and implement more efficient and effective cybersecurity practices. Practically, XAI application helps to create a more cooperative interaction between AI systems and human professionals, which increases the effectiveness of the whole system.

5.4 Challenges and Limitations

XAI can change the cybersecurity systems, yet the challenges related to the integration of XAI to cybersecurity systems are multiple. One of the negative aspects is the cost of calculations related to the application of the methods of XAI, in particular, the use of complex models, including SHAP and LIME, which require substantial processing resources and time to explain. Also, there is a risk of data privacy as XAI systems usually need to access sensitive information to produce relevant explicates. The problem of balancing between transparency and sensitive information protection is crucial. Moreover, it may be difficult and expensive to implement XAI into current cybersecurity systems because these systems may need to be changed or upgraded to support the features of explainability. Such issues should be resolved to achieve the mass acceptance and use of XAI in cybersecurity.

5.5 Recommendations

In a bid to apply XAI effectively to cybersecurity, it is advisable that the professionals focus on the choice of the model depending on the performance specifications of the system and trade-off between interpretability and computational efficiency. Other AI models such as decision trees and rule systems are simpler and easier to interpret but less detailed than a more complex AI model. Training of XAI models regularly is necessary in order to make sure that they can be adjusted to changing threats and still be explainable. The continuous validation of these models should as well be done to determine how effective they are in the real world and also to adjust them to the best performance. Also, the best practices of data privacy should be applied by the cybersecurity teams to make sure that the explanations that XAI systems give do not jeopardize confidential information.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

VI. CONCLUSION

6.1 Summary of Key Points

The paper has discussed how Explainable AI (XAI) can be used to improve cybersecurity by increasing the level of transparency and trust in automated systems of threat detection. The key results are that the XAI enables decipherable meaning of AI-driven decisions, which enable cybersecurity experts to justify and comprehend the activities of the system. This helps to build trust in AI much more, as black-box systems are less relied upon, and decisions are made much faster during security breaches. XAI can also improve the effectiveness of threat detection by facilitating a better understanding of interactions between human specialists and AI models, which in turn fosters confidence among security staff, which subsequently results in more effective and reliable cybersecurity practices. The study revealed that XAI systems like LIME, SHAP, and decision trees have the potential to narrow the divide between the complicated AI and human users so that automated systems could be useful, responsible, and trustworthy even when it comes to cybersecurity.

6.2 Future Directions

Subsequent studies in XAI need to address the development of the interpretability methods to work with more advanced, high-performance models applied in cybersecurity, including deep learning-based systems. With more complex AI models, it is necessary to develop new methods to interpret decisions of this type of system with accuracy and a level of understanding by cybersecurity experts. Furthermore, the development of new uses of XAI in uncharted fields of cybersecurity, including threat hunting, risk assessment, and proactive protection systems, is likely to enhance the operation of security greatly. The study of balancing between explainability and computational efficiency as well as data privacy will be instrumental in the extensive use of XAI. Moreover, the combination of XAI and real-time threat detection systems and the creation of hybrid models that balance interpretability and high detection rates are all avenues through which the field can be further advanced in the future.

REFERENCES

[1] Azmi, S. K. (2025). Enhancing Java Virtual Machine Performance for Scalable Artificial Intelligence and Machine Learning Workloads. Well Testing Journal, 34(S3), 566-580.
[2] Syed Khundmir Azmi. (2025). Enhancing Java Virtual Machine Performance for Scalable Artificial Intelligence and Machine Learning Workloads. Well Testing Journal, 34(S3), 566–580. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/221
[3] Azmi, S. K. (2025). Enhancing Java Virtual Machine performance for scalable artificial intelligence and machine learning workloads. Well Testing Journal, 34(S3), 566–580.
[4] Azmi, S. K. (2025). LLM-Aware Static Analysis: Adapting Program Analysis to Mixed Human/AI Codebases at Scale. Global Journal of Engineering and Technology Advances, 24(03), 260-269.
[5] Syed, Khundmir Azmi. (2025). LLM-Aware Static Analysis: Adapting Program Analysis to Mixed Human/AI Codebases at Scale. Global Journal of Engineering and Technology Advances. 24. 10.30574/gjeta.2025.24.3.0284.
[6] Azmi, S. K. (2025). LLM-aware static analysis: Adapting program analysis to mixed human/AI codebases at scale. Global Journal of Engineering and Technology Advances, 24(3), 260–269.
[7] Azmi, Syed Khundmir. "LLM-Aware Static Analysis: Adapting Program Analysis to Mixed Human/AI Codebases at Scale." Global Journal of Engineering and Technology Advances, vol. 24, no. 3, 30 Sept. 2025, pp. 260–269, https://doi.org/10.30574/gjeta.2025.24.3.0284. Accessed 7 Oct. 2025.
[8] Azmi, S. K. (2023). Trust but Verify: Benchmarks for Hallucination, Vulnerability, and Style Drift in AI-Generated Code Reviews. Well Testing Journal, 32(1), 76-90.
[9] Syed Khundmir Azmi. (2023). Trust but Verify: Benchmarks for Hallucination, Vulnerability, and Style Drift in AI-Generated Code Reviews. Well Testing Journal, 32(1), 76–90. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/229
[10] Azmi, S. K. (2023, February 6). Trust but verify: Benchmarks for hallucination, vulnerability, and style drift in AI-generated code reviews. Well Testing Journal, 32(1), 76–90.
[11] Syed, Khundmir Azmi. (2023). Secure DevOps with AI-Enhanced Monitoring. International Journal of Science and Research Archive. 9. 10.30574/ijsra.2023.9.2.0569.
[12] Syed, Khundmir Azmi. "Secure DevOps with AI-Enhanced Monitoring." International Journal of Science and Research Archive, vol. 9, no. 2, 30 June 2023, pp. 1193–1200, https://doi.org/10.30574/ijsra.2023.9.2.0569. Accessed 13 Oct. 2025.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

[13] Azmi, S. K. (2022). From Assistants to Agents: Evaluating Autonomous LLM Agents in Real-World DevOps Pipeline. Well Testing Journal, 31(2), 118-133.
[14] Azmi, S. K. (2022). From assistants to agents: Evaluating autonomous LLM agents in real-world DevOps pipeline. Well Testing Journal, 31(2), 118–133.
[15] Syed Khundmir Azmi. (2022). From Assistants to Agents: Evaluating Autonomous LLM Agents in Real-World DevOps Pipeline. Well Testing Journal, 31(2), 118–133. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/230
[16] Azmi, S. K. (2022). Green CI/CD: Carbon-Aware Build & Test Scheduling for Large Monorepos. Well Testing Journal, 31(1), 199-213.
[17] Syed Khundmir Azmi. (2022). Green CI/CD: Carbon-Aware Build & Test Scheduling for Large Monorepos. Well Testing Journal, 31(1), 199–213. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/231
[18] Azmi, S. K. (2022). Green CI/CD: Carbon-aware build & test scheduling for large monorepos. Well Testing Journal, 31(1), 199–213.
[19] Azmi, S. K. (2021). Computational Yoshino-Ori Folding for Secure Code Isolation in Serverless It Architectures. Well Testing Journal, 30(2), 81-95.
[20] Azmi, S. K. (2021, October 28). Computational Yoshino-Ori folding for secure code isolation in serverless IT architectures. Well Testing Journal, 30(2), 81–95.
[21] Syed Khundmir Azmi. (2021). Computational Yoshino-Ori Folding for Secure Code Isolation in Serverless It Architectures. Well Testing Journal, 30(2), 81–95. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/237
[22] Azmi, S. K. (2021). Riemannian Flow Analysis for Secure Software Dependency Resolution in Microservices Architectures. Well Testing Journal, 30(2), 66-80.
[23] Azmi, S. K. (2021). Riemannian flow analysis for secure software dependency resolution in microservices architectures. Well Testing Journal, 30(2), 66–80.
[24] Syed Khundmir Azmi. (2021). Riemannian Flow Analysis for Secure Software Dependency Resolution in Microservices Architectures. Well Testing Journal, 30(2), 66–80. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/236
[25] Azmi, S. K. (2025). Voronoi partitioning for secure zone isolation in software-defined cyber perimeters. Global Journal of Engineering and Technology Advances, 24(03), 431-441.
[26] Azmi, S. K. (2025). Voronoi partitioning for secure zone isolation in software-defined cyber perimeters. Global Journal of Engineering and Technology Advances, 24(3), 431–441
[27] Syed, Khundmir Azmi. (2025). Voronoi partitioning for secure zone isolation in software-defined cyber perimeters. Global Journal of Engineering and Technology Advances. 24. 431-441. 10.30574/gjeta.2025.24.3.0294.
[28] Azmi, Syed Khundmir. "Voronoi Partitioning for Secure Zone Isolation in Software-Defined Cyber Perimeters." Global Journal of Engineering and Technology Advances, vol. 24, no. 3, 30 Sept. 2025, pp. 431–441, https://doi.org/10.30574/gjeta.2025.24.3.0294. Accessed 13 Oct. 2025.
[29] Syed, Khundmir Azmi. (2025). Zero-Trust Architectures Integrated With Blockchain For Secure Multi-Party Computation In Decentralized Finance. INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS. 13. 2320-2882
[30] Syed, Khundmir Azmi. (2025). Bott-Cher Cohomology For Modeling Secure Software Update Cascades In Iot Networks. INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS. 13. g1-g12.
[31] Azmi, S. K. (2025). Bott-Cher Cohomology for Modeling Secure Software Update Cascades in IoT Networks. International Journal of Creative Research Thoughts (IJCRT), 13(9)
[32] Syed, Khundmir Azmi. (2025). Retrieval-Augmented Requirements: Using RAG To Elicit, Trace, And Validate Requirements From Enterprise Knowledge Bases.
[33] Azmi, S. K. (2025, September 9). Retrieval-Augmented Requirements: Using RAG to Elicit, Trace, and Validate Requirements from Enterprise Knowledge Bases. International Journal of Creative Research Thoughts (IJCRT), 13(9).
[34] Syed, Khundmir Azmi. (2025). Hypergraph-Based Data Sharding for Scalable Blockchain Storage in Enterprise IT Systems. Journal of Emerging Technologies and Innovative Research. 12. g475-g487.
[35] Azmi, S. K. (2025). Kirigami-Inspired Data Sharding for Secure Distributed Data Processing in Cloud Environments. JETIR, 12(4).
[36] Syed, Khundmir Azmi. (2025). Kirigami-Inspired Data Sharding for Secure Distributed Data Processing in Cloud Environments. Journal of Emerging Technologies and Innovative Research. 12. o78-o91.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

[37] Syed, Khundmir Azmi. (2024). Human-in-the-Loop Pair Programming with AI: A Multi-Org Field Study across Seniority Levels. International Journal of Innovative Research in Science Engineering and Technology. 13. 20896-20905. 10.15680/IJIRSET.2024.1312210|.

[38] Azmi, S. K. (2024, October). Klein bottle-inspired network segmentation for untraceable data flows in secure IT systems. IRE Journals. https://www.irejournals.com/formatedpaper/1711014.pdf

[39] Syed, Khundmir Azmi & Azmi,. (2024). Klein Bottle-Inspired Network Segmentation for Untraceable Data Flows in Secure IT Systems. 8. 852-862.

[40] Syed, Khundmir Azmi & Azmi,. (2023). Quantum Zeno Effect for Secure Randomization in Software Cryptographic Primitives. 7. 2456-8880.

[41] Azmi, S. K. (2024, March). Quantum Zeno effect for secure randomization in software cryptographic primitives. IRE Journals. Retrieved from https://www.irejournals.com/paper-details/1711015

[42] Azmi, S. K. (2024). Cryptographic hashing beyond SHA: Designing collision-resistant, quantum-resilient hash functions. International Journal of Science and Research Archive, 12(2), 3119–3127.

[43] Syed, Khundmir Azmi. (2024). Cryptographic Hashing Beyond SHA: Designing collision-resistant, quantum-resilient hash functions. International Journal of Science and Research Archive. 13. 3119-3127. 10.30574/ijsra.2024.12.2.1238.

[44] Azmi, Syed Khundmir. "Cryptographic Hashing beyond SHA: Designing Collision-Resistant, Quantum-Resilient Hash Functions." International Journal of Science and Research Archive, vol. 12, no. 2, 31 July 2024, pp. 3119–3127, https://doi.org/10.30574/ijsra.2024.12.2.1238. Accessed 9 Oct. 2025.

[45] Syed Khundmir Azmi. (2023). Photonic Reservior Computing or Real-Time Malware Detection in Encrypted Network Traffic. Well Testing Journal, 32(2), 207–223. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/244

[46] Azmi, S. K. (2023, August 31). Photonic reservoir computing or real-time malware detection in encrypted network traffic. Well Testing Journal, 32(2), 207–223.

[47] Azmi, S. K. (2023). Photonic Reservior Computing or Real-Time Malware Detection in Encrypted Network Traffic. Well Testing Journal, 32(2), 207-223.

[48] Syed, Khundmir Azmi. (2025). Algebraic geometry in cryptography: Secure post-quantum schemes using isogenies and elliptic curves. International Journal of Science and Research Archive. 10. 1509-1517. 10.30574/ijsra.2023.10.2.0965.

[49] Azmi, Syed Khundmir. "Algebraic Geometry in Cryptography: Secure Post-Quantum Schemes Using Isogenies and Elliptic Curves." International Journal of Science and Research Archive, vol. 10, no. 2, 31 Dec. 2023, pp. 1509–1517, https://doi.org/10.30574/ijsra.2023.10.2.0965. Accessed 15 Oct. 2025.

[50] Azmi, S. K. (2023). Algebraic geometry in cryptography: Secure post-quantum schemes using isogenies and elliptic curves. IJSRA. https://ijsra.net/sites/default/files/IJSRA-2023-0965.pdf

[51] Syed, Khundmir Azmi. (2022). Bayesian Nonparametrics in Computer Science: Scalable Inference for Dynamic, Unbounded, and Streaming Data. 5. 399-407.

[52] Azmi, S. K. (2022, April). Bayesian nonparametrics in computer science: Scalable inference for dynamic, unbounded, and streaming data. IRE Journals. https://www.irejournals.com/formatedpaper/1711044.pdf

[53] Syed Khundmir Azmi. (2022). Computational Knot Theory for Deadlock-Free Process Scheduling in Distributed IT Systems. Well Testing Journal, 31(1), 224–239. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/243

[54] Azmi, S. K. (2022, March 30). Computational knot theory for deadlock-free process scheduling in distributed IT systems. Well Testing Journal, 31(1), 224–239.

[55] Azmi, S. K. (2021, September). Markov Decision Processes with Formal Verification: Mathematical Guarantees for Safe Reinforcement Learning. IRE Journals, 5(3) https://www.irejournals.com/formatedpaper/1711043.pdf

[56] Syed, Khundmir Azmi. (2021). Markov Decision Processes with Formal Verification: Mathematical Guarantees for Safe Reinforcement Learning. 5. 418-428.

[57] Conde Camillo da Silva, R., Oliveira Camargo, M. P., Sanches Quessada, M., Lopes, A. C., Diassala Monteiro Ernesto, J., & Pontara da Costa, K. A. (2022). An Intrusion Detection System for Web-Based Attacks Using IBM Watson. *IEEE Latin America Transactions*, 20(2), 191-197. https://doi.org/10.1109/TLA.2022.9661457

[58] Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2020). A historical perspective of explainable Artificial Intelligence. *WIREs Data Mining and Knowledge Discovery*, 11(1). https://doi.org/10.1002/widm.1391

[59] Dwivedi, R., Dave, D., Naik, H., Singhal, S., Rana, O., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2022). Explainable AI (XAI): Core Ideas, Techniques and Solutions. *ACM Computing Surveys*, 55(9), 1–33. https://doi.org/10.1145/3561048

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 8, Issue 5, September – October 2025 ||

DOI: 10.15680/IJCTECE.2025.0805012

[60] Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

[61] Mia, L. (2025). Evaluating the Trade-offs Between Explainability and Security in AI-Powered Cyber Defense. https://doi.org/10.2139/ssrn.5140427

[62] Moustafa, N., Koroniotis, N., Keshk, M., Zomaya, A. Y., & Tari, Z. (2023). Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions. *IEEE Communications Surveys & Tutorials*, 25(3), 1775-1807. https://doi.org/10.1109/COMST.2023.3280465

[63] Niteen, N., & Kurian, S. M. (2023). Exploring Explainable AI, Security and Beyond: A Comprehensive Review. *International Journal on Emerging Research Areas*, 3(2). https://ijera.in/index.php/IJERA/article/view/5

[64] Srivastava, G., Jhaveri, R. H., Bhattacharya, S., Pandya, S., Rajeswari, Maddikunta, P. K. R., Yenduri, G., Hall, J. G., Alazab, M., & Gadekallu, T. R. (2022). XAI for Cybersecurity: State of the Art, Challenges, Open Issues and Future Directions. *ArXiv:2206.03585 [Cs]*. https://arxiv.org/abs/2206.03585

[65] Jimmy, F. (2021). Emerging Threats: The Latest Cybersecurity Risks and the Role of Artificial Intelligence in Enhancing Cybersecurity Defenses. *Valley International Journal Digital Library*, 9(2), 564–574. https://doi.org/10.18535/ijsrm/v9i2.ec01