NJESR/August-2023/Volume-5/Issue-8 E-ISSN-2582-5836
DOI-10.53571/NJESR.2023.5.8.61-68

Explainable Quantum Generative AI For Cyber Threat Simulation And Defense

Sabu.K.J
Lecturer
Department Of Computer Engineering
Government Polytechnic College
Kothamangalam
Kerala

(Received:20July2023/Revised:5August2023/Accepted:17August2023/Published:31August2023)

Abstract

The issue of understanding and explaining how AI makes decisions has become more important as artificial intelligence (AI) is used more widely in modern cyber threat intelligence (CTI) systems. Explainable Quantum Generative AI for Cyber Threat Simulation and Defence uses Generative AI (QGAI) powered by quantum technology to understand and predict cyber threats, and then uses these insights to build stronger defenses. Since quantum computing is better at solving complex problems than traditional systems, QGAI can create very realistic attack simulations and find new weaknesses that regular methods might miss. The "explainable" part is important for trust and checking the system's fairness because it helps security teams look for biases and understand how the AI reaches its conclusions, especially since quantum systems often work like a "black box". While machine learning models are powerful at finding complex and changing cyber threats, they are hard to understand and can make people lose confidence in the results. This study looks at using Explainable Artificial Intelligence (XAI), which includes tools like attention-based visualisations and SHAP (SHapley Additive exPlanations), to improve how we interpret CTI systems. AI is changing how we protect against cyber threats and keep digital assets safe. This innovative technology is setting new standards for cybersecurity resilience, from finding threats early to automatically responding to attacks. The study also discusses challenges like the mental effort needed to process information and risks from malicious attacks and suggests future work in areas like rules, governance, and policies. According to my research, being able to explain how AI works is not just helpful it's a must for building strong and trustworthy cyber defence systems.

Keywords:- Explainable AI (XAI), Cybersecurity, Interpretable Machine Learning, Threat Detection, Intrusion Detection Systems (IDS)

Introduction

The growing complexity, sophistication, and scale of cyber attacks have made artificial intelligence (AI) a key part of modern cyber defense. AI-based solutions are important for protecting digital systems because they offer features such as adaptive threat modeling, automatic response systems, and real-time detection of unusual activity. Many of these systems work as "black boxes," meaning they don't provide much insight into how they make decisions, especially those that use deep learning and complex models[1]. In areas where trust, accountability, and following rules are very important, this lack of transparency can be a big problem. AI works by constantly improving and creating new data outputs on its own. Generative AI is a big change from traditional methods. It uses advanced deep learning techniques like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). These tools help create high-dimensional data by manipulating hidden layers and using probability-based models. GANs use two networks generator and discriminator that work against each other to improve the quality of the output over time. VAEs encode data into simpler forms, from which new data can be made[2]. These models don't just create new data; they often generate original content like high-quality images or natural language that's almost impossible to tell apart from what humans create. Generative AI is being used in different areas, each using its unique ability to create things on its own. In creative fields, it's being used to automatically produce art, music, and text, which challenges ideas about human creativity and who owns the content. In healthcare, it's speeding up drug discovery by creating new molecule designs and helping with medical imaging to improve diagnosis[1-2]. In cybersecurity, it's being used for detecting threats and simulating attacks, which helps improve both defensive and offensive strategies. However, the more we depend on generative AI, the more challenges it brings. Ethical issues are a big concern, especially with deepfakes and biased algorithms. Deepfake technology, which uses GANs, can create very realistic fake videos and audio, which can damage the trust in information and harm public confidence. Also, if the training data has biases, it can lead to unfair results in decision-making systems, making social inequalities worse. From a security angle, generative AI can introduce new ways for attacks to happen. It can create code or design complicated phishing attacks, making cyber threats bigger and more complex. These threats are made worse when generative AI is used

to spread fake news quickly, making it harder to detect and stop[3]. Privacy is another major concern, especially when it comes to using personal data to train these models. The large datasets needed for training often include private information, raising questions about who owns the data, whether people have given permission, and if data can be re-identified even if it's anonymized. These developments are putting pressure on the current laws and rules around AI, including the General Data Protection Regulation (GDPR) and other privacy laws. Explainable AI (XAI) aims to fix this by making AI systems more transparent and easier to understand for people who use them. While XAI is becoming more popular in areas like healthcare and finance, it's not being used much in cybersecurity, and when it is, it's not very consistent[4]. Cybersecurity professionals often have to make fast, important decisions that need clear explanations for why AI systems give certain alerts or labels. If they can't see how the AI made its decision, they might ignore a real threat or trust a wrong result. This paper looks at the role of XAI in changing how we defend against cyber attacks. I review different explainability methods, check if they work well for security situations, and suggest a new framework that fits into Security Operations Centers (SOCs). I argue that explainability isn't just a nice-to-have feature; it's a must-have for making sure AI is trustworthy and effective in defending against cyber threats[5].

The Need For Explainability In Cyber Defense

AI is being used in cybersecurity to improve how we find and handle threats, but there's a big problem with understanding how these AI systems make their decisions. In important areas like national defense, key infrastructure, and banking, it's not just helpful it's necessary to be able to understand and trust what AI systems are telling us. Explainable AI, or XAI, helps by giving clear and easy-to-understand reasons for the decisions AI makes. This allows cybersecurity experts to check, challenge, or take action on alerts in a confident way[4-5]. From a legal point of view, rules like the European Union's GDPR say people have the right to know why automated systems, including those in cybersecurity, make certain decisions [4]. The U.S. National Institute of Standards and Technology also highlights explainability as a key part of its AI Risk Management Framework [5]. Cybersecurity analysts are often faced with too many false alarms from detection tools, which leads to alert fatigue. When AI decisions lack clear explanations, it makes things worse, causing distrust or not using advanced tools properly [6]. Having explainable AI not only speeds up responses by showing why an alert is raised, but it also helps people and machines work better together, making the overall security stronger. Hackers are also using AI to trick systems

through attacks like evasion and poisoning. Interpretable AI models let security teams spot weaknesses and make their threat detection systems more reliable [7]. In this situation, explainability serves as both a technical and strategic advantage, helping organizations stay strong, open, and responsible in a dangerous online world.

Gen AI In Cybersecurity Defense

Generative AI, which is a part of artificial intelligence, is changing the way cybersecurity works by providing better ways to spot, study, and deal with security threats. Unlike regular AI, which mainly looks for patterns and odd behavior, Generative AI creates new data, tests possible attacks, and quickly learns from new dangers. A report from IDC says that by 2025, 60% of the world's top 2000 companies will use AI-based security tools to fight growing cyber attacks[7]. This move towards smarter security is also supported by Gartner, which predicts that more than 50% of organizations will use AI-powered services in their security efforts by 2024. The fast use and growth of AI in cybersecurity show a big change towards more flexible and smart security systems, meeting the changing needs of today's digital world[7-8].

Benefits Of Generative AI In Cybersecurity

Generative AI is leading the way in modern cybersecurity, providing many benefits that make traditional security systems better. Here are some of the main advantages[7-9]:

1. **Better Threat Detection** Generative AI can look at huge amounts of data to find patterns and strange behavior that might signal a threat.These systems keep learning and improving over time, making them more accurate. This means they can spot threats quicker and more reliably than humans, who might get overwhelmed by the amount of data[7].
2. **Automatic Response** One big benefit of Generative AI is that it can automatically respond to cyber threats.

Using set rules and real-time data, AI can quickly take actions like cutting off infected devices or blocking bad IP addresses. This fast reaction helps stop attackers before they can do much damage[7].

3. **Predicting Future Threats** Generative AI is great at predicting possible threats by looking at past data.

This helps organizations take action before an attack happens, lowering the chance of a

successful breach. It also helps in planning resources better, making sure cybersecurity efforts are focused on the most important areas[8].

4. **More Efficient Use of Resources** Generative AI handles routine tasks like monitoring and analyzing data, allowing human teams to focus on more complex and important work.This smart use of resources improves the overall security of an organization[8].
5. **Better Fraud Detection** Generative AI models can find unusual patterns that might show fraud. By learning from new data, these models stay up-to-date with new fraud methods, offering strong protection against things like identity theft and financial scams. This ongoing learning keeps fraud detection effective and reliable[9].

XAI In Security Applications: Use Cases And Techniques

Using Explainable AI (XAI) in cybersecurity helps security experts better understand how AI systems assess threats. This part looks at important situations where XAI is used, such as spotting intrusions, identifying malware, and recognizing phishing attempts, and explains the XAI methods used in each case[9].

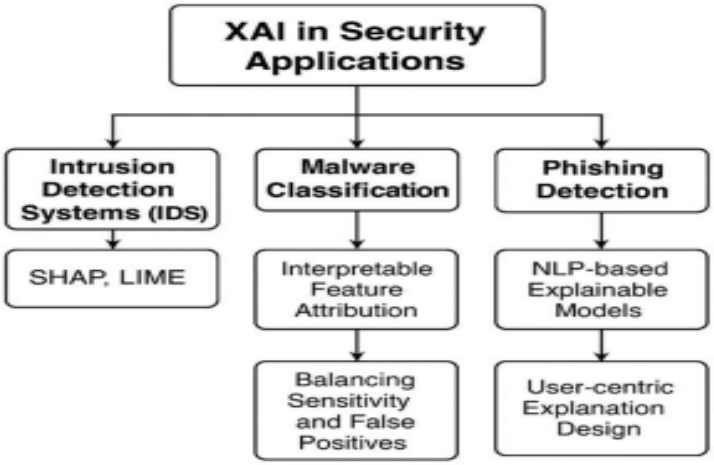

Figure 1: XAI Security Applications

Intrusion Detection Systems (IDS)

Intrusion Detection Systems benefit a lot from being able to explain their decisions, as analysts need to review many alerts with little background information. Methods like SHAP and LIME, which are part of explainable AI (XAI), help explain how a model reached its decision in IDS. These methods show which network features, like unusual port activity or packet rates, were most important in labeling an event as malicious [10].

Studies show that using these feature explanations boosts analyst confidence and speeds up the process of checking false positives [10].

Malware Classification

Deep learning models used for malware detection often don't provide clear explanations, making it hard to trust their results. XAI techniques such as gradient-based saliency maps and layer-wise relevance propagation (LRP) are used to show which parts of the data, like certain byte sequences or API calls, had the biggest impact on a malware prediction [10]. These insights help reverse engineers and forensic analysts understand how the model reached its conclusions and spot new attack methods.

Phishing And Social Engineering Detection

Phishing detection increasingly uses natural language processing (NLP) models to analyze emails and websites. Model-agnostic XAI methods can point out specific words or phrases, like urgent language or misspellings, that helped classify something as phishing [1-4]. This transparency helps train end users and allows security teams to create more effective responses. These uses show how XAI serves two important roles: improving detection accuracy and helping people work together with AI. By making model outputs clearer, security teams can better focus on risks, check out strange activities, and adjust their defenses[10].

Proposed Framework For XAI-Enabled

To make Explainable AI (XAI) work in cybersecurity, I suggest a modular framework that adds easy-to-understand machine learning parts into current security tools like Security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR) systems. This setup helps provide clear, useful, and fast threat information while keeping the system flexible and able to grow as needed[10].

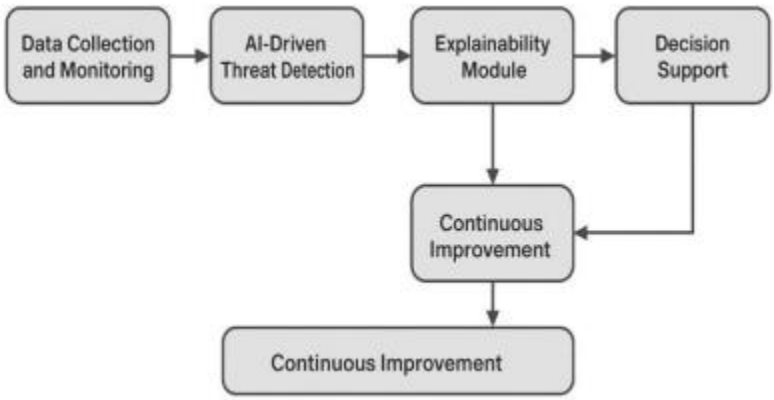

Figure 2:. Framework for XAI-Enabled Cyber Defense

Conclusion

Generative AI is definitely changing how cybersecurity works, helping create smarter, more flexible, and more effective ways to spot and deal with threats. Using generative AI in cybersecurity brings a lot of benefits, like better threat detection, stronger defense systems, and faster responses. As cyber threats get more complicated and more common, it's important to use Explainable AI (XAI) in cybersecurity. This isn't just about improving how well threats are found it's also about making things clearer, more trustworthy, and easier for people to understand. This article looked at the importance of being able to explain AI decisions in security, reviewed the latest XAI methods, and studied how they work in areas like detecting intrusions, classifying malware, and spotting phishing attempts. I suggested a framework that puts users first for adding explainability into cybersecurity systems and pointed out challenges like making these systems work well at scale, not overwhelming users with too much information, and dealing with attempts to trick the system. Putting Explainable AI into Cyber Threat Intelligence systems is a big step forward in making cybersecurity more trusted, clear, and effective. As cyber threats become more complex and numerous, AI-powered CTI systems provide a powerful way to find, sort, and predict harmful activities quickly and accurately.

References

[1].Tanikonda, A., Pandey, B. K., Peddinti, S. R., & Katragadda, S. R. (2022). Advanced AI-driven cybersecurity solutions for proactive threat detection and response in complex ecosystems. Journal of Science & Technology, 3(1).

[2]. Paul, E. M., Stanley, U. M., Kessie, J. D., & Dolapo, M. (2023). Adversarial machine learning in cybersecurity: Mitigating evolving threats in AI-powered defense systems.

[3]. Rjoub, G., Bentahar, J., Wahab, O. A., Mizouni, R., Song, A., Cohen, R., ... & Mourad, A. (2023). A survey on explainable artificial intelligence for cybersecurity. IEEE Transactions on Network and Service Management, 20(4), 5115-5140.

[4]. Kalejaiye, A. N. (2022). REINFORCEMENT LEARNING-DRIVEN CYBER DEFENSE FRAMEWORKS: AUTONOMOUS DECISION-MAKING FOR DYNAMIC RISK PREDICTION AND ADAPTIVE THREAT RESPONSE STRATEGIES. International Journal of Engineering Technology Research & Management (IJETRM), 6(12), 92-111.

[5]. Kehoe, A., Wittek, P., Xue, Y., & Pozas-Kerstjens, A. (2021). Defence against adversarial attacks using classical and quantum-enhanced Boltzmann machines. Machine Learning: Science and Technology, 2(4), 045006.

[6]. Kehoe, A., Wittek, P., Xue, Y., & Pozas-Kerstjens, A. (2021). Defence against adversarial attacks using classical and quantum-enhanced Boltzmann machines. Machine Learning: Science and Technology, 2(4), 045006.

[7]. Alfurhood, B. S., Mankame, D. P., Dwivedi, M., & Jindal, N. (2023). Artificial Intelligence and Cybersecurity: Innovations, Threats, and Defense Strategies. Journal of Advanced Zoology, 44(S2), 4715-4721.

[8]. Suryotrisongko, H., Musashi, Y., Tsuneda, A., & Sugitani, K. (2022). Robust botnet DGA detection: Blending XAI and OSINT for cyber threat intelligence sharing. IEEE Access, 10, 34613-34624.

[9]. Sreevallabh Chivukula, A., Yang, X., Liu, B., Liu, W., & Zhou, W. (2022). Game theoretical adversarial deep learning. In Adversarial Machine Learning: Attack Surfaces, Defence Mechanisms, Learning Theories in Artificial Intelligence (pp. 73-149). Cham: Springer International Publishing.

[10]. Jaladi, D. S., & Vutla, S. (2022). Medical Decision-Making with the Help of Quantum Computing and Machine Learning: An In-Depth Analysis. International Journal of Acta Informatica, 1(1), 199-215.