RADemics

Combining Explainable AI and Advanced Visual Analytics for Threat Attribution and Response Justification in Machine Learning Frameworks

Rachhpal Singh, Dineshkumar A
PUNJABI UNIVERSITY, SRI RANGANATHAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

Combining Explainable AI and Advanced Visual Analytics for Threat Attribution and Response Justification in Machine Learning Frameworks

[1]Rachhpal Singh, Assistant Professor, Department of Computer Science, Punjabi University, Patiala, Punjab, India. rachhpal@pbi.ac.in

[2]Dineshkumar A, Assistant Professor, Department of Artificial Intelligence and Data Science, Sri Ranganathar Institute of Engineering and Technology, Tamil Nadu, India. dineshkumarannadurai207@gmail.com

Abstract

This chapter explores the integration of Explainable AI (XAI) and Advanced Visual Analytics (AVA) for enhancing cybersecurity, with a particular focus on threat attribution and response justification in machine learning frameworks. As the complexity of cyber threats increases, the need for transparent, interpretable, and justifiable automated decision-making systems becomes critical. This work delves into the role of XAI in providing transparency and trust in machine learning models, while AVA offers intuitive visualizations for real-time monitoring and decision support. By combining these technologies, organizations can achieve a deeper understanding of security events, improve incident response, and ensure compliance with legal and ethical standards. The chapter highlights key advancements, presents relevant case studies, and discusses the legal, regulatory, and technical challenges of integrating XAI and AVA in cybersecurity. The synergy between XAI and AVA was positioned as a crucial factor in evolving cybersecurity frameworks to meet modern threats effectively.

Keywords: Explainable AI, Advanced Visual Analytics, Threat Attribution, Response Justification, Cybersecurity, Machine Learning.

Introduction

The rapid evolution of digital infrastructure and the growing sophistication of cyber threats have introduced new challenges in safeguarding critical systems and sensitive data [1]. Traditional cybersecurity approaches, while effective to some extent, often struggle to keep up with the speed and complexity of modern attacks [2-4]. Cybercriminals are increasingly using advanced tactics, techniques, and procedures (TTPs) to breach defenses, making it essential for organizations to adopt more proactive and intelligent approaches [5-8]. In response, machine learning (ML) and artificial intelligence (AI) technologies are being integrated into cybersecurity frameworks to detect, analyze, and mitigate threats [9]. However, these technologies, especially when deployed in automated decision-making systems, require mechanisms to ensure their actions are understandable, accountable, and justifiable [10,11]. Without transparency, the effectiveness of AI and ML models in cybersecurity can be compromised, leading to potential mistrust or incorrect responses to incidents [12].

Explainable AI (XAI) plays a pivotal role in addressing the challenges of transparency and trust in AI-driven decision-making systems [13]. In cybersecurity, XAI ensures that the rationale behind automated security actions was clear and understandable to human analysts, providing much-needed accountability [14]. XAI techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), enable security experts to understand why specific decisions, such as flagging a potential threat, were made by machine learning models [15,16]. This transparency was crucial not only for validating the effectiveness of the model but also for regulatory compliance [17,18]. As cybersecurity models become increasingly complex, XAI ensures that these systems can be trusted, audited, and corrected when necessary, enhancing their overall reliability and usability [19].

Advanced Visual Analytics (AVA) has emerged as a powerful tool for enhancing cybersecurity decision-making, particularly in real-time threat detection and monitoring [20]. By converting complex data from security systems into intuitive visual representations, AVA makes it easier for analysts to interpret vast amounts of data and identify potential threats faster [21,22]. Interactive dashboards, heatmaps, and threat visualizations enable security teams to understand the scope and context of security events, facilitating quicker and more accurate responses [23]. When combined with AI-driven insights, AVA enhances the ability to track and respond to evolving threats, offering a comprehensive overview of the attack landscape [24,25]. The integration of AVA into cybersecurity workflows ensures that analysts are equipped with the tools to not only detect threats but also understand their implications in real-time, driving more effective decision-making.