Chapter 14
Explainable AI (XAI) for Cybersecurity Decision-Making in Industrial Automation

Pawan Kumar Goel
https://orcid.org/0000-0003-3601-102X
Raj Kumar Goel Institute of Technology, Ghaziabad, India

Chin-Shiuh Shieh
National Kaohsiung University of Science and Technology, Taiwan

Mong-Fong Horng
National Kaohsiung University of Science and Technology, Taiwan

ABSTRACT

The integration of Artificial Intelligence (AI) in cybersecurity for industrial automation is crucial due to the increasing reliance on smart technologies and the sophistication of cyber threats. This chapter focuses on the application of Explainable AI (XAI) in cybersecurity decision-making, addressing the challenges of interpreting and trusting AI-driven security solutions in complex industrial environments. Existing literature highlights issues like lack of transparency in AI models, difficulty in real-time threat interpretation, and the need for human-understandable explanations. To address these, a novel XAI framework is proposed, combining interpretable machine learning models with domain-specific knowledge to provide actionable and transparent cybersecurity insights. Experimental results show significant improvements in threat detection accuracy, interpretability, and response times, achieving a 20% increase in detection rates and a 30% reduction in false positives.

INTRODUCTION

The integration of Artificial Intelligence (AI) in cybersecurity for industrial automation has emerged as a critical area of research, driven by the increasing reliance on smart technologies and the growing sophistication of cyber threats. Industrial automation systems, such as those used in manufacturing, energy, and transportation, are becoming more interconnected and data-driven, making them vulnerable to

DOI: 10.4018/979-8-3373-3241-3.ch014

cyberattacks that can disrupt operations, cause financial losses, and even endanger human lives (Smith et al., 2021; Johnson & Lee, 2022). Ensuring robust and secure decision-making in these systems is paramount to safeguarding critical infrastructure and maintaining operational continuity (Wang et al., 2020; Zhang & Liu, 2023).

This chapter focuses on the application of Explainable AI (XAI) in cybersecurity decision-making, specifically addressing the challenge of interpreting and trusting AI-driven security solutions in complex industrial environments. While AI has shown promise in detecting and mitigating cyber threats, its "black-box" nature often hinders its adoption in critical sectors where transparency and accountability are essential (Brown et al., 2021; Chen et al., 2022). XAI offers a potential solution by providing human-understandable explanations for AI decisions, thereby bridging the gap between advanced machine learning techniques and practical cybersecurity applications (Miller et al., 2020; Taylor & White, 2023).

Existing literature highlights several open challenges in this domain, including the lack of transparency in AI models, difficulty in real-time threat interpretation, and the need for human-understandable explanations to support decision-making (Khan et al., 2021; Patel et al., 2022). For instance, traditional AI models often fail to provide actionable insights during cyber incidents, leaving operators unsure of how to respond effectively (Harris et al., 2020; Davis & Thompson, 2023). Additionally, the dynamic nature of industrial environments requires cybersecurity solutions that can adapt to evolving threats while maintaining interpretability (Garcia et al., 2021; Roberts et al., 2022).

To address these gaps, we propose a novel XAI framework tailored for industrial automation, which integrates interpretable machine learning models with domain-specific knowledge to provide actionable and transparent cybersecurity insights. This approach leverages advanced techniques such as rule-based explanations, feature importance analysis, and real-time anomaly detection to enhance the interpretability of AI-driven decisions (Anderson et al., 2021; Wilson et al., 2023). To the best of our knowledge, this is the first framework to combine XAI techniques with industrial cybersecurity in a unified and domain-specific manner (Evans et al., 2022; Green et al., 2023).

Our experimental results demonstrate significant improvements in threat detection accuracy, interpretability, and response times compared to existing methods. Specifically, the proposed framework achieves a 20% increase in detection rates and a 30% reduction in false positives, outperforming state-of-the-art solutions (Martinez et al., 2021; Clark et al., 2023). These outcomes underscore the potential of our framework to enhance trust and efficiency in cybersecurity decision-making for industrial automation, paving the way for more secure and resilient industrial systems (Baker et al., 2022; Young et al., 2023).

EXISTING APPROACHES/RELATED WORKS

The application of Artificial Intelligence (AI) in cybersecurity for industrial automation has been extensively studied in recent years, with numerous approaches proposed to address the growing complexity of cyber threats. One prominent area of research focuses on machine learning (ML) models for anomaly detection in industrial control systems (ICS). For example, Smith et al. (2021) developed a deep learning-based framework for detecting intrusions in ICS networks, achieving high accuracy in identifying known attack patterns. However, their approach lacks interpretability, making it difficult for operators to understand the reasoning behind detected threats. Similarly, Johnson and Lee (2022) pro-