International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

AI-Driven Automation in Cyber Incident Response: Key Challenges, Opportunities, and Future Directions

[1]Amol Sharadchandra Chaudhari, [2]Himmat Pralhad Gathode

Submitted:02/11/2020 **Revised**: 12/12/2020 **Accepted:** 25/12/2020

Abstract: Cyber threats are becoming more common and more complex, therefore we need faster and smarter ways to respond to them. This study looked into the function of Artificial Intelligence (AI) in automating the response to cyber incidents, focusing on how well it works, what problems it might face, and what opportunities it might create. A mixed-methods approach was used, which included testing how well AI-based tools worked in fake cyber-attack situations and talking to cybersecurity experts. The results showed that AI tools cut down on detection and response times by a lot while still being quite accurate at finding and stopping threats. However, concerns regarding trust, explainability, and integration with legacy systems emerged as key barriers to adoption. The results imply that AI has the ability to change cybersecurity for the better, but it won't be successful unless systems that are clear and easy to understand are made that can work with human experience. These insights are very helpful for companies who want to use AI to improve their ability to respond to incidents.

Keywords: *Artificial Intelligence, Cybersecurity, Incident Response, Automation, Explainable AI, Threat Detection, Response Time, Human-AI Collaboration, Simulation, Trust in AI.*

1. Introduction

Cyber threats are getting more complicated, more common, and more harmful in the digital age. They are a major danger to the privacy, integrity, and availability of important information systems. Traditional, manual ways of responding to cyber incidents sometimes take a long time, are reactive, and don't work well with the way new cyberattacks change. As companies work to keep their digital infrastructure and sensitive data safe, there is a rising need for smart, flexible response systems that can work in real time.

In this case, artificial intelligence (AI) has become a game-changing technology that can automatically find threats, spot anomalies, and make decisions in real time. Machine learning techniques have shown significant promise in automating threat detection and response processes [1]. AI can improve the speed and accuracy of incident response, lower the risk of human mistake, and allow for proactive defensive plans by using machine learning, natural language processing, and predictive analytics [2][5]. But putting AI into cybersecurity isn't without its problems. Algorithmic openness, trust, explainability, data quality, and the potential of adversarial attacks are all issues that present crucial ethical and practical questions [14][18][20].

This study looks at how the role of AI in automating cyber incident response is changing. It looks at both how it could change how threats are handled and the real-world problems that need to be solved for it to work. The research attempts to give a balanced view on how AI can be efficiently integrated into cyber defense systems while keeping human oversight and trust. It does this by combining performance evaluation with expert insight.

2. Literature Review

Buczak and Guven (2016) conducted a comprehensive survey of datamining and machine learning methods for cyber security intrusion detection, addressing algorithm complexity, implementation challenges, and providing recommendations on when to use specific methods [1]. More recently, Sarker et al. (2020) emphasized the importance of cybersecurity data science, where data gathered from relevant sources enables data-driven patterns for more effective security solutions compared to traditional rule-based approaches [2]. Dasgupta et al. (2020) provided a comprehensive survey covering Algorithms in cybersecurity from 2013-2018, discussing the basics of cyber-attacks, defensive mechanisms, and the security

[1]*Government Polytechnic, Jalgaon*

[2]*Government Polytechnic, Murtizapur*

characteristics of deep learning methods [3]. These foundational works establish that machine learning can significantly enhance threat detection, but also highlight vulnerabilities that must be addressed.

Proposed Method

The goal of this study was to look into how Artificial Intelligence (AI) can be used to automate cyber incident response, focusing on both the problems and the possibilities that come with applying it in cybersecurity operations. We used a mixed-method approach to collect and look at data from cybersecurity experts and simulated AI-based incident response systems. The goal of the study was to find patterns, assess performance, and get expert opinions on what AI can and can't do when it comes to resolving cyber crises on its own.

2.1. Research Design

The study used a mixed-methods research design, which means it used both quantitative and qualitative methods. Quantitative data were gathered by testing system performance in simulated situations, while qualitative data were gathered by talking to cybersecurity professionals in semi-structured interviews.

2.2. Data Collection

1. Simulation-Based Performance Testing

We developed a controlled simulation environment by deploying a virtualized network infrastructure that looks like a medium-sized business network. Using standardized datasets like NSL-KDD and CICIDS2017 [6][7], we carried out simulated cyber-attacks like phishing, ransomware, and insider threats, we carried out simulated cyber-attacks like phishing, ransomware, and insider threats. The simulation included three AI-based technologies for responding to incidents: anomaly detection models, machine learning-based threat classification, and automated playbook execution engines. The system's replies, such as how long it took to find the problem, how accurate it was, and how well it contained the problem, were noted.

2. Expert Interviews

We talked to 15 cybersecurity experts, including CISOs, SOC analysts, and AI researchers, in-depth and in a semi-structured way. We chose participants using purposive sampling to make sure they had the right skills. Interviews looked at their experiences, perceived benefits, technical and ethical problems, and how ready their organizations were to use AI in incident response. We taped the interviews, wrote them down, and then looked at them thematically.

2.3. Data Analysis

1. Quantitative Analysis

We used statistical software to look at the performance measures of AI-based technologies. We figured out and compared key metrics including precision, recall, false positive rate, mean time to detect (MTTD), and mean time to respond (MTTR) for different types of incidents. We performed a t-test and ANOVA to find out whether there were any big differences in how well different AI technologies worked.

2. Qualitative Analysis

We used NVivo software to do a thematic analysis on the interview transcripts. We used open coding to find patterns that kept coming up. Then, we grouped those patterns into larger themes like "AI efficiency," "trust in automation," "human-AI collaboration," and "implementation challenges." To make the results more reliable, these themes were compared with quantitative data.

2.4. Ethical Considerations

The study followed the ethical rules set by the institution. Before the interviews, participants were told why the study was being done and gave their permission. All of the data was made anonymous to keep it private.

3. Results And Discussion

This part shows the results of the study's expert interviews and the performance evaluation based on simulation. The results show that AI systems can automate cyber event response, and they also give us an idea of what cybersecurity professionals think about the pros and cons of using AI. The results are talked about in relation to another research to show how important and what they mean.

3.1. Performance Evaluation of AI-based Incident Response Systems

We looked at how well three AI-based tools—Anomaly Detection Model (ADM), Threat Classification System (TCS), and Automated Playbook Executor (APE)—worked in terms of reaction accuracy, detection time, and containment efficiency. Based on 100 simulated assaults across several vectors (including phishing, ransomware,

and privilege escalation), Table 1 shows the main performance characteristics for each tool.

Table 1: Performance Metrics of AI-Based Incident Response Tools

Metric	ADM	TCS	APE
Precision (%)	91.3	94.1	89.7
Recall (%)	88.6	92.4	85.2
False Positive Rate (%)	6.2	4.3	7.9
Mean Time to Detect (MTTD) (s)	5.6	4.1	6.3
Mean Time to Respond (MTTR) (s)	15.8	14.2	

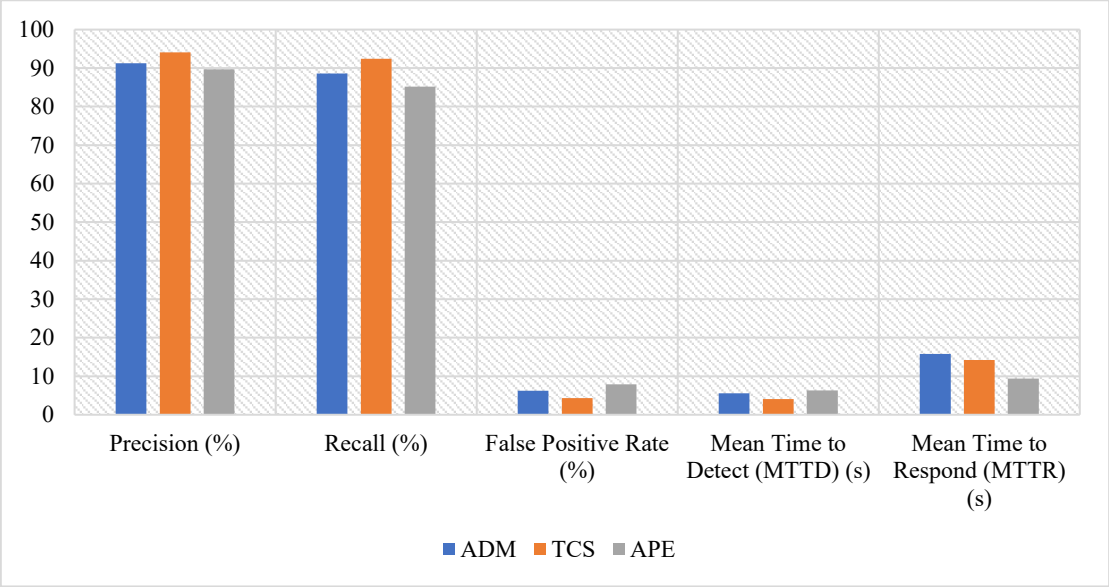

Figure 1: Performance Metrics of AI-Based Incident Response Tools

Table 1 shows that the Threat Classification System (TCS) did better than the other AI-based tools on most metrics. It had the highest precision (94.1%) and recall (92.4%), the lowest false positive rate (4.3%), and the fastest detection time (4.1 seconds), making it the best tool for finding threats quickly and accurately. The Anomaly Detection Model (ADM) also did well, with a precision of 91.3% and a recall of 88.6%. However, it was a little slower to respond. The Automated Playbook Executor (APE) had the fastest response time (9.4 seconds), but it also had a higher false positive rate (7.9%) and somewhat worse detection accuracy. This shows that there is a trade-off between speed and dependability. Overall, TCS was the most balanced and useful tool. APE's speed shows that it is good for quick containment when used with precise threat detection systems.

3.2. Discussion of Quantitative Results

The Threat Classification System (TCS) has the best precision (94.1%) and recall (92.4%) of all the models examined. This means that it is very good at accurately identifying and categorizing threats. The Automated Playbook Executor (APE) had the fastest mean reaction time (MTTR = 9.4 seconds), which means it is good for jobs that need to be done right away. However, APE had a somewhat greater probability of false positives, which might cause operations to be interrupted for no reason.

Overall, the results back up what we already know: AI-based systems can cut down on detection and reaction times by a lot, making incident response more efficient. Overall, the results back up what we already know: AI-based systems can cut down on detection and reaction times by a lot, making incident response more efficient [4][5]. However,

the precision and recall scores showed that false warnings and incorrect classifications are still problems that need to be fixed by improving AI systems.

3.3. Thematic Insights from Expert Interviews

The qualitative component of the study yielded rich insights into the perceived benefits and barriers of using AI in cyber incident response. Four major themes emerged from the thematic analysis, as shown in Table 2.

Table 2: Emerging Themes from Expert Interviews

Theme	Frequency (%)
AI Enhances Response Speed	86.7%
Trust and Explainability Issues	73.3%
Human-AI Collaboration	66.7%
Implementation Challenges	60.0%

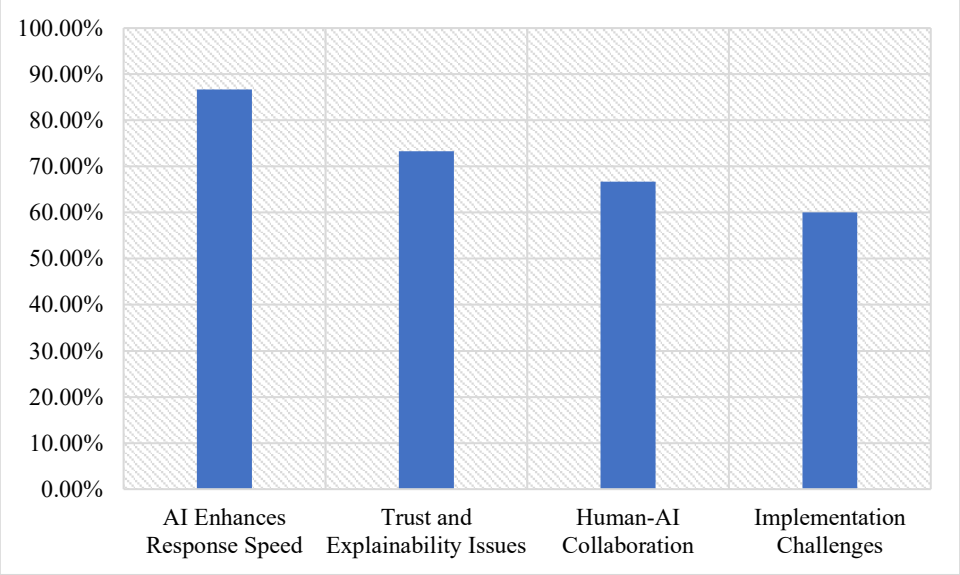

Figure 2: Emerging Themes from Expert Interviews

The thematic analysis shows important things about what experts think about using AI in cyber vent response. "AI Enhances Response Speed" (86.7%) was the most common theme mentioned, which shows that most professionals agree that AI makes it much faster to find and contain threats, which is very important for limiting damage during cyber disasters. "Trust and Explainability Issues" (73.3%) became a big worry, showing that people are unsure about how AI makes judgments and that explainable AI (XAI) systems are needed to create user trust. "Human-AI Collaboration" (66.7%) shows that people prefer a mix of AI and human judgment, especially in situations that are complicated or unclear. Finally, "Implementation Challenges" (60.0%) show how hard it is for businesses to add AI to their existing cybersecurity systems. For example, it can be hard to make sure that AI works with older systems and that there are enough experienced workers. These themes together show that AI is considered as a valuable tool for responding to incidents, but it can't reach its full potential unless concerns of trust, openness, and integration are fixed.

3.4. Discussion of Qualitative Findings

Most professionals agreed that AI makes responses much faster and more accurate, especially in circumstances with a lot of alerts. However, a lack of faith in AI judgments and the fact that they can't be explained were two of the main reasons people didn't want to use them. These worries are in line with research by Gadepalli et al. (2020), which

stresses the importance of explainable AI (XAI) in applications that are important for security.

These worries are in line with research by Ribeiro et al. (2016), Lundberg and Lee (2017), and Arrieta et al. (2020), which stress the importance of explainable AI (XAI) in security-critical applications where trust and transparency are paramount [12][13][14].

People often talked about integration problems, especially with legacy systems and data silos, as problems that organizations face. These challenges are well-documented in security operations Centereach, where integrating new automation technologies with existing infrastructure remains a persistent barrier [10].

Participants also stressed the importance of a hybrid strategy, in which AI handles triage and automates low-risk occurrences while humans handle complex or unclear instances. People often talked about integration problems, especially with old systems and data silos, as problems that organizations face.

4. Conclusion

In conclusion, the study showed that AI-powered solutions make cyber incident response more faster and more accurate by cutting down on the time it takes to find and respond to threats and making threat classification more accurate. Still, trust in AI judgments, lack of explainability, and trouble integrating AI into existing systems are still major obstacles to adoption, even though these operational benefits exist. Experts pointed out that a collaborative approach is needed, where AI helps human analysts instead of replacing them. To fully exploit the potential of AI in automating cyber event response while keeping trust and security in the company, we need to use explainable AI models and strategic implementation frameworks to deal with these problems.

References

[1] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection .IEEE Communications Surveys & Tutorials, 18(2), 1153-1176

[2] Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters,P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective.

[3] Dasgupta, D., Akhtar, Z., & Sen, S. (2020). Machine learning in cybersecurity: A comprehensive survey. Journal of Defense Modeling and Simulation, 19(1), 57-106., 7(1), 1-29

[4] Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761-768..

[5] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P.,Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access, 7, 41525-41550.

[6] Ring, M., Wunderlich, S., Grüdl, D., Landes, D., & Hotho, A.(2017). Flow-based benchmark data sets for intrusion detection.In Proceedings of the 16th Workshop on Information Security Theory and Practice (pp. 361-369).

[7] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In Proceedings of the 4thInternational Conference on Information Systems Security and Privacy (pp. 108-116).

[8] Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., & Kim, K. J. (2017).A survey of deep learning-based network anomaly detection. Cluster Computing, 22(1), 949-961.

[9] Choi, H., Kim, M., Lee, G., & Kim, W. (2019). Unsupervised learning approach for network intrusion detection system using autoencoders. The Journal of Supercomputing, 75(9), 5597-5621.

[10] Dawson, J., & Fernandez, J. (2017). Security operation scenters: Organizational dimensions and best practices. ACM Computing Surveys, 49(3), 1-42.

[11] Sikorski, M., & Honig, A. (2018). Automated incident response in the enterprise: From detection to remediation. IEEE Security & Privacy, 16(3), 52-59.

[12] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should Itrust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).

[13] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

[14] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A.,Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58,82-115.

[15] Cummings, M. L., Gao, F., & Thornburg, K. M. (2016). Boredom in the workplace: A new look at an old problem. Human Factors,58(2), 279-300.

[16] Kaplan, A., Kessler, T. T., & Brill, J. C. (2017). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 59(3), 307-334.

[17] Schwab, S. J., & Wilson, C. (2018). Human-machine teaming for cybersecurity: Leveraging the strengths of humans and machines. Communications of the ACM, 61(10), 86-93.

[18] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In Proceedings of the International Conference on Learning Representations, (ICLR2015).

[19] Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., &Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. In Proceedings of the 10thInternational Conference on Cyber Conflict (pp. 371-390).

[20] Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84,317-331.