

Explainable and Responsible AI Security Tools for Trust, Accountability, and Compliance

Abu Sarwar Zamani

 <https://orcid.org/0000-0002-0135-2337>

Prince Sattam bin Abdulaziz University, Saudi Arabia

Aisha Hassan Abdalla Hashim

 <https://orcid.org/0000-0001-6331-1373>

International Islamic University Malaysia, Malaysia

Anwer Mustafa Hlilal

Prince Sattam bin Abdulaziz University, Saudi Arabia

Vishal Barot

 <https://orcid.org/0000-0002-9139-9541>

Gujarat Technological University, India

Durgesh Nandan

 <https://orcid.org/0000-0002-9762-3559>

SR University, India

Table of Contents

Preface	xvii
Acknowledgment	xxvi
Introduction	xxvii
 Chapter 1	
Foundations of Explainable and Responsible AI in Cybersecurity	1
<i>Irfan Ahmad Khan, Rai University, Ahmedabad, India</i>	
<i>Faizan Samdani, Waed Khadmat Al Munzal Le Cygne Commercial</i>	
<i>Center, Saudi Arabia</i>	
 Chapter 2	
Explainability in Malware and Ransomware Detection	29
<i>Ketan Sarvakar, Ganpat University, India</i>	
<i>Dolly Prajapati, Ganpat University, India</i>	
 Chapter 3	
Explainable AI Techniques for Intrusion Detection Systems	67
<i>Rashid Ayub, King Saud University, Saudi Arabia</i>	
<i>Abu Sarwar Zamani, Prince Sattam bin Abdulaziz University, Saudi</i>	
<i>Arabia</i>	
<i>Vishal Girishbhai Barot, GTU Institute of Technology and Research,</i>	
<i>India</i>	
<i>Sumbul Alam, University of Bisha, Saudi Arabia</i>	
 Chapter 4	
Adversarial Machine Learning and Robust Explainability	97
<i>M. Premkumar, SSM Institute of Engineering and Technology, India</i>	
<i>R. Sudha, PSNA College of Engineering and Technology, India</i>	

Chapter 5

Fairness, Bias, and Ethical Risks in AI-Based Cyber Defense: Towards Responsible and Trustworthy Cybersecurity	131
--	-----

Rashid Ayub, King Saud University, Saudi Arabia

Ayesha Ayub, Asia Pacific University, Malaysia

Mohammed Bilal Ashraf, Asia Pacific University, Malaysia

Mohammed Rayan, BITS Pilani, Dubai, UAE

Chapter 6

Human-Centered AI and Decision Support in Security Operations Centers	153
--	-----

Ankit Jain, Pranveer Institute of Technology, Kanpur, India

Anita Shukla, Pranveer Singh Institute of Technology, Kanpur, India

Puspraj Singh Chauhan, Indian Institute of Technology, Indore, India

Imran Ullah Khan, Integral University, Lucknow, India

Chapter 7

Governance, Compliance, and Regulatory Frameworks for Responsible AI ...	181
--	-----

Yogesh Mehta, Shree Guru Gobind Singh Tricentenary University, India

Durgesh Nandan, SR University, India

Chapter 8

Explainable AI in Emerging Cybersecurity Domains	211
--	-----

Subuhi Kashif Ansari, Jazan University, Saudi Arabia

Yasir Ahmad, Jazan University, Saudi Arabia

Nithinsha Shajahan, Jazan University, Saudi Arabia

Sadia Husain Javed, Jazan University, Saudi Arabia

Anjali Appukuttan, Jazan University, Saudi Arabia

Anam Shahzad, Netaji Subhas University of Technology, India

Chapter 9

Large Language Models in Cybersecurity: Explainability, Trust, and Responsible Deployment	249
---	-----

Deepak Sharma, Gujarat Technological University-Institute of Technology and Research, India

Nancy Girdhar, School of Engineering, Computing, and Mathematics, University of Plymouth, UK

Ankit Modi, Gujarat Technological University-Institute of Technology and Research, India

Chapter 10

Evaluation Metrics and Benchmarking for Explainable AI in Cybersecurity: XAI-Guard - Explainable Deep Learning Framework for Trustworthy Threat Detection and Cybersecurity.....	271
--	-----

*Sultan Ahmad, Prince Sattam Bin Abdulaziz University, Al Kharj, Saudi
Arabia & Chandigarh University, Gharuan, India*

Shaik Khaja Mohiddin, Siddhartha Academy of Higher Education, India

*Shaik Sharmila, Vignan's Nirula Institute of Technology and Science for
Women, India*

*Vijaya Chandra Jadala, School of Computer Science and Artificial
Intelligence, SR University, India*

*Hikmat A. M. Abdeljaber, Faculty of Information Technology, Applied
Science Private University, Jordan*

Chapter 11

Visualization and Interpretability Dashboards for Security Analysts	299
---	-----

Nidhi Bhavsar, Thakur College of Engineering and Technology, India

Komal Dhule, Thakur College of Engineering and Technology, India

*Apeksha Waghmare, Thakur College of Engineering and Technology,
India*

Kajal Patel, Thakur College of Engineering and Technology, India

Kriti Mitra, Thakur College of Engineering and Technology, India

Chapter 12

Federated and Distributed Learning for Secure and Explainable Cyber Defense	335
--	-----

Vishal Barot, GTU - Institute of Technology and Research, India

Abu Sarwar Zamani, INTI International University, Malaysia

Ankit Modi, GTU - Institute of Technology and Research, India

Vivek Patel, GTU - Institute of Technology and Research, India

Dipak Dabhi, GTU - Institute of Technology and Research, India

Vivekkumar Prasad, Independent Researcher, India

Chapter 13

Future Directions Towards Trustworthy Autonomous and Self-Explainable Cyber Defense Systems	369
--	-----

*Shital Amit Karande, Bharati Vidyapeeth's College of Engineering for
Women, India*

*Bindu R. Garg, College of Engineering, Bharati Vidyapeeth (Deemed to
be University), Pune, India*

*Sultan Ahmad, College of Computer Engineering and Sciences,
Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia
& University Center for Research and Development (UCRD),
Department of Computer Science and Engineering, Chandigarh
University, Gharuan, India*

*Rajesh Prasad, College of Engineering, Bharati Vidyapeeth (Deemed to
be University), Pune, India*

Chapter 14

Case Studies and Real-World Deployments of Explainable AI in Cybersecurity	399
---	-----

Sachi Bhavsar, SAL Engineering and Technical Institute, India

Madhuri Parekh, SAL Engineering and Technical Institute, India

Shivani Doshi, SAL Engineering and Technical Institute, India

Krupa Patel, SAL Engineering and Technical Institute, India

Meeta Matnani, SAL Engineering and Technical Institute, India

Shalini Chaturvedi, SAL Engineering and Technical Institute, India

Compilation of References	437
--	------------

About the Contributors	479
-------------------------------------	------------

Index.....	487
-------------------	------------

Chapter 6

Human–Centered AI and Decision Support in Security Operations Centers

Ankit Jain

 <http://orcid.org/0000-0002-6687-5518>

Pranveer Institute of Technology, Kanpur, India

Anita Shukla

 <http://orcid.org/0000-0002-6687-5518>

Pranveer Singh Institute of Technology, Kanpur, India

Puspraj Singh Chauhan

Indian Institute of Technology, Indore, India

Imran Ullah Khan

Integral University, Lucknow, India

ABSTRACT

The growing complexity of cyber threats has made Security Operations Centers (SOCs) essential for real-time threat detection, incident response, and infrastructure protection. However, traditional black-box AI models often reduce trust and transparency in cybersecurity decisions. This chapter explores human-centered Explainable Artificial Intelligence (XAI) for improving interpretable threat detection, anomaly analysis, risk prioritization, and analyst decision-making. It also presents an AIoT-based “Explainable Intelligent SOC Monitoring System for Smart Enterprise Networks” integrating IoT network sensors, edge devices, AI intrusion

DOI: 10.4018/979-8-2600-1970-2.ch006

Copyright © 2027, IGI Global Scientific Publishing. Copying or distributing in print or electronic forms without written permission of IGI Global Scientific Publishing is prohibited. Use of this chapter to train generative artificial intelligence (AI) technologies is expressly prohibited. The publisher reserves all rights to license its use for generative AI training and machine learning model development.

detection, SHAP and LIME explainability, federated learning, and adaptive alert prioritization. A real-time SOC dashboard provides attack explanations, threat scores, and response support. Human–AI collaboration, analyst feedback, fairness, bias mitigation, trust, accountability, and regulatory concerns are also discussed to support more transparent, reliable, and trustworthy cyber defense systems.

1. INTRODUCTION TO HUMAN-CENTERED CYBERSECURITY

A human-centered approach to cyber security puts people, situation awareness, and decision accountability at the forefront of cyber defense activities. The modern Security Operations Center (SOC) generates immense amounts of alerts related to endpoints, cloud services, identities, industrial assets, and Internet of Things (IoT) devices. Machine learning technologies assist with the analysis of those volumes, but the utility of machine learning hinges upon human capability to interpret results. An alert without an explanation can make people feel even more hesitant when it comes to responding to a threat in a time-sensitive environment. A human-centered approach to machine learning tries to decrease fatigue among analysts, improve the quality of triaging processes, and produce data that can be used to make justified decisions (Arrieta et al., 2020; Adadi & Berrada, 2018).

1.1 Need for Trustworthy Security Automation

Alert fatigue, understaffing, and attack campaigns with better coordination are some of the challenges faced by security analysts today. Automation is inevitable as reviewing every alert manually becomes difficult. However, over-reliance on automated processes can result in blind spots for security analysts when the rationale behind a model's decision that a certain event was malicious remains unknown. Reliable automation would require accurate output results, explanations of how the results were reached, indicators of confidence levels of those results, and opportunities for feedback. It would need to allow analysts to understand why an event was scored a certain risk level and help them identify the most critical cases from background noise (Ali et al., 2023; Ribeiro et al., 2016). Figure 1. illustrates a human-centered artificial intelligence (AI) workflow for security operations centers, integrating network sensors, cloud logs, AI detection, explainability, analyst decision support, response actions, and feedback learning.