

Chapter 2

The Role of Explainable AI (XAI) in Forensic Investigations

Yara Shamoo

 <https://orcid.org/0009-0004-2459-0807>

Saint Leo University, USA

ABSTRACT

The increasing reliance on Artificial Intelligence (AI) in digital forensic investigations has brought forward significant advancements in the identification, analysis, and interpretation of digital evidence. However, one of the major challenges remains the “black-box” nature of many AI models, which often hinder trust, accountability, and transparency. Explainable AI (XAI) emerges as a critical solution, providing clarity into the decision-making processes of AI systems. In the context of digital forensics, XAI facilitates investigators’ understanding of AI-driven tools, ensuring that conclusions drawn from automated analyses are interpretable, verifiable, and legally admissible. This chapter explores the role of XAI in digital forensics, highlighting its potential to enhance the reliability and transparency of AI-based forensic systems, while addressing key challenges and ethical considerations. We examine the integration of XAI in forensic workflows, discuss various techniques for explainability, and evaluate real-world case studies where XAI has contributed to improved forensic outcomes.

DOI: 10.4018/979-8-3373-0857-9.ch002

1. INTRODUCTION

Explainable Artificial Intelligence (XAI) refers to a set of processes, techniques, and methods that make the decision-making process of AI models understandable to humans. In contrast to traditional “black-box” AI models, which often operate in ways that are not easily interpretable, XAI aims to provide transparency, accountability, and interpretability for complex AI systems. By making AI’s reasoning transparent, stakeholders—including investigators, policymakers, and end-users—can gain trust in the system’s decisions and ensure they are making fair, responsible, and ethical choices. In the realm of digital forensics, where the need for transparency and reliability is paramount, the application of XAI can significantly enhance the effectiveness and accountability of forensic investigations.

AI has increasingly become a core component in modern digital forensic investigations, aiding in the collection, analysis, and interpretation of digital evidence. However, the inherent complexity and opacity of many AI techniques, especially deep learning models, pose challenges in forensics, where decisions need to be explained and validated. The inability to explain why or how an AI system made a particular decision can lead to trust issues, especially in high-stakes legal and security contexts (Hall, Sakzad, & Choo, 2022). Thus, XAI has become crucial in ensuring that AI’s role in digital forensics is both effective and justifiable.

The core advantage of XAI lies in its ability to offer decision-making transparency. It enables forensic analysts to understand the rationale behind an AI model’s predictions, which is vital in legal contexts where the findings of forensic investigations can be challenged in court. In digital forensics, where evidence can be easily contested, it is critical that AI systems offer explanations for their actions in a manner that is accessible and understandable to non-experts.

Moreover, XAI can help uncover biases that may influence AI predictions. This is of particular concern in forensics, as any unintended bias in the data or the AI model could lead to incorrect conclusions, such as misidentifying a suspect or overlooking a critical piece of evidence. By allowing forensic investigators to trace and understand AI’s decision-making process, XAI can prevent these types of errors (Costantini, De Gasperis, & Olivieri, 2019).

As the field of AI and digital forensics continues to evolve, the integration of XAI techniques into forensic workflows will be critical. Investigators will be able to leverage AI for tasks such as anomaly detection, pattern recognition, and evidence authentication while ensuring that the results remain explainable and verifiable. The combination of AI’s powerful analytical capabilities and XAI’s transparency could lead to more effective, equitable, and defensible forensic processes.

30 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/the-role-of-explainable-ai-xai-in-forensic-investigations/367310

Related Content

Enhanced YOLOv11 for Image-Based Anomaly Detection in Freight Train Gate Chains

Han Jianfeng, Guoqing Cui, Jinnan Du, Chao Wang and Liying Zhang (2025). *International Journal of Intelligent Information Technologies* (pp. 1-18).

www.irma-international.org/article/enhanced-yolov11-for-image-based-anomaly-detection-in-freight-train-gate-chains/386138

Inclusive Finance in the Digital Age: New Insights on Fintech's Role in Financial Inclusion for Low-Literacy Populations

Yuan Yuan, Ainul Huda Jamiland Nurlaili A. B. Ghani (2026). *Regulation, Emerging Risk, and Ethics in FinTech and AI* (pp. 197-224).

www.irma-international.org/chapter/inclusive-finance-in-the-digital-age/388601

Multi-Agent Simulation Collision Avoidance of Complex System: Application to Evacuation Crowd Behavior

Mohammed Chennoufi, Fatima Bendella and Maroua Bouzid (2018). *International Journal of Ambient Computing and Intelligence* (pp. 43-59).

www.irma-international.org/article/multi-agent-simulation-collision-avoidance-of-complex-system/190632

AI and Social Networks: A Transformative Partnership for Education

Ken Lui (2025). *AI Deployment and Adoption in Public Administration and Organizations* (pp. 165-196).

www.irma-international.org/chapter/ai-and-social-networks/373918

A Lane Identifying Approach of the Intelligent Vehicle in Complex Condition: Intelligent Vehicle in Complex Condition

Botao Wu and Huijuan Wang (2019). *International Journal of Ambient Computing and Intelligence* (pp. 25-44).

www.irma-international.org/article/a-lane-identifying-approach-of-the-intelligent-vehicle-in-complex-condition/238052