International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Exploring the Dangers of AI a review on Malicious Use and Abuse in Crime

Devika Devaraj. G. D [a], *Dr. Keerthi Kumar . H. M* [b]

Research Scholar [a]*, Malnad College of Engineering, Hassan, Karnataka-573201, India*
Associate Professor[b], *Malnad College of Engineering, Hassan, Karnataka-573201, India*

ABSTRACT

The paper presents a thorough analysis of AI's dual role in cyber security, outlining both its benefits and vulnerabilities. It categorizes AI-related crime into two main types: AI as a tool for crime and AI as a target for crime. The escalation of cyber attacks due to AI enhancements and the potential for physical harm from AI-related crimes are highlighted. Investigating AI-related crimes poses significant challenges, necessitating innovative strategies in AI forensics. The urgency of interdisciplinary collaboration to address these threats is emphasized, stressing the need for comprehensive preventive measures. Overall, the paper underscores the critical importance of understanding and mitigating the risks associated with the malevolent use of AI.

Keywords: Artificial intelligence, AI crime, AI forensics, security threats, malicious AI¸ AI tools.

1. Introduction

The paper addresses the security threats and crimes associated with artificial intelligence and the need for AI forensics. It emphasizes the risks introduced by artificial intelligence algorithms and training data, as well as the concern about cyber attacks enhanced by artificial intelligence. Additionally, the paper highlights the potential for AI-related crime that can physically harm individuals. It proposes a taxonomy for AI crime, categorizing it into AI as a tool crime and AI as a target crime. The challenges of investigating AI crime and the importance of developing novel strategies in AI forensics are also discussed. The document provides a comprehensive review on AI security threats, AI-related crime, and cybercrime, underscoring the significance of addressing and preventing the malicious use of AI through collaborative research efforts. The paper delves into AI-related security threats and crimes, advocating for the development of AI forensics to address them. It proposes a taxonomy for categorizing AI crime and stresses the importance of collaborative efforts to prevent malicious AI use.

2. Literature Survey

The literature review covers AI security threats, AI crime, and digital forensics, employing a comprehensive search across diverse sources. It highlights challenges in handling multimedia data and addressing issues of irreproducibility in AI systems. Additionally, it discusses the relevance of similarity analysis in digital forensics, emphasizing the importance of forensic stakeholders acquiring expertise in AI to tackle emerging challenges effectively.

3. Related work

An overview of research on AI security threats, AI-related crime, and digital forensics from various perspectives. It mentions that the term "AI crime" was initially introduced in the humanities field, highlighting security threats and malicious uses of AI that can lead to various crimes. The section also delves into digital forensics, defining it as the use of scientifically derived methods for the collection and analysis of digital evidence. It emphasizes the importance of adhering to principles such as meaning, errors, transparency, trustworthiness, reproducibility, and experience in the forensic process. Additionally, the document discusses the challenges and solutions presented in forensic sub-fields like smart phone, cloud, and IoT forensics. Moreover, the paper addresses the unique challenges posed by AI forensics, particularly in investigating AI crimes. It proposes future research directions for AI forensics, including AI exploration, similarity analysis, adversarial attack detection, and damage assessment. The AI exploration aspect involves identifying how AI is utilized in crimes and understanding the intentions behind the development and application of AI systems in criminal activities. Furthermore, the document highlights the need for forensic investigators to develop expertise in AI to effectively address the challenges in AI forensic. We gain insights into the evolving landscape of AI security threats, AI crime, digital forensics, and the emerging field of AI forensics. This comprehensive overview sets the stage for understanding the complexities and implications of AI-related crimes and the critical need for innovative forensic strategies to address them effectively.

Based on the literature review, this article defines the term AI crime and classifies it into two categories: AI as a tool crime and AI as a target crime, inspired by a taxonomy of cybercrime: Computer as a tool crime and Computer as a target crime.

1. AI as a Tool Crime: AI as a tool crime refers to crimes where AI technology is used as a tool to commit an offense. Examples include AI-driven phishing attacks, AI-powered malware, and AI-generated fake content, among others.

2. AI as a Target Crime: AI as a target crime refers to crimes where AI systems are the primary target. Examples include adversarial attacks on AI systems, data poisoning, and AI model theft, among others.

3. Foreseeable AI Crimes and Forensic Techniques: Through the proposed taxonomy, foreseeable AI crimes are systematically studied, and related forensic techniques are addressed.

1. 1 AI as a Tool Crime:

Foreseeable AI crimes under this category include:

- AI-driven Phishing Attacks
- AI powered Malware
- AI-generated Fake Content
- AI-enhanced Social Engineering Attacks

Forensic Techniques:

- Behavioral Analysis of AI Algorithms
- Pattern Recognition in AI-generated Content
- AI Attribution Techniques

2.1 AI as a Target Crime:

Foreseeable AI crimes under this category include:

- Adversarial Attacks on AI Systems
- Data Poisoning
- AI Model Theft

Forensic Techniques:

- Adversarial Robustness Testing
- Data Provenance Analysis
- Model Watermarking and Authentication

4. Methodology

The methodology in the context of artificial intelligence forensics involves adapting traditional forensic principles to the unique challenges posed by AI systems. This includes ensuring the meaning of evidence remains unchanged, minimizing errors in the forensic process, establishing transparency and trustworthiness in the forensic methods employed, striving for reproducibility of results, and developing expertise in the field of AI crime and forensics.

Moreover, given the complexity and scale of AI systems, the forensic process needs to be adjusted to accommodate the vast amount of data involved in AI investigations. Particularly, with AI systems focusing on multimedia data such as images or sound, which are already challenging in traditional digital forensics, a new approach to evidence collection and analysis is necessary. In dealing with the unpredictability of AI algorithms and the potential for adversarial attacks, a proactive stance on preventing and detecting such attacks is crucial. This may involve developing defense methods to classify adversarial examples correctly, or alternatively, focusing on the detection of adversarial examples through statistical methods or additional neural networks.

Ultimately, reframing the methodology in AI forensics requires a holistic approach that addresses the unique characteristics of AI systems, the challenges posed by AI-related crimes, and the need for expertise in both traditional forensic techniques and AI technologies.

This figure serves to visually represent the differences between these two models within an AI system.

AI Crime		Techniques	Related Research
Advanced Computer as Tool Crime		Crime Al chatbot , Deep fake, etc	[3] , [4] , [6] , [7]
Advanced Computer as Target Crime		Social engineering, Vulnerability scanner, etc	[9] ,[12]
Physical Crime		Drone swarm, hardware attack etc	[4]
Training system attack		Data modification	[5]
Training system theft	Inference system cracking	Vulnerability attacks used in computer as target Crime	[10] , [11]
Inference System Abuse		White-box attack, black-box attack	[13]

5. Result

The paper discusses the security threats and crimes associated with artificial intelligence and the need for AI forensics. The highlight the risks posed by AI algorithms and training data, as well as the concern about cyber attacks enhanced by AI. They propose a taxonomy for AI crime, categorizing it into AI as tool crime and AI as target crime. The paper emphasizes the challenges in investigating AI crime and the importance of addressing malicious AI use through research. The paper also address the issue of reproducibility in AI forensics, noting the difficulty in collecting all elements of an AI system due to technical and legal issues. Limited collection of evidence makes it challenging to reproduce past states of AI systems for investigation . Overall, the paper provides insights into the complexities of AI security threats, AI-related crimes, and the need for innovative approaches in AI forensics to combat malicious AI use.

6. Discussion

While AI has revolutionized various industries and brought about significant progress, it has also introduced complex risks and challenges that need to be addressed urgently .One crucial aspect to consider is the intricate issue of collecting evidence related to AI crimes. Due to technical and legal constraints, forensic stakeholders face difficulties in gathering all elements of an AI system, such as the training system, learning model, dataset, trained model, and inference system. This limited collection of evidence poses a significant obstacle for investigators, as the reproducibility of AI systems is a key principle in digital forensics. Moreover, the large-scale nature of AI systems, particularly those focusing on multimedia data like images and sounds, presents a unique challenge for traditional forensic tools and processes. Furthermore, expertise in AI forensics is paramount for forensic stakeholders to effectively address the challenges posed by AI crimes. Understanding AI systems, structures, and environments is essential for forensic investigators to propose viable solutions and techniques for AI forensic investigations .Overall, a comprehensive and collaborative approach involving researchers, policymakers, AI professionals, and forensic examiners is essential to tackle the security threats, crimes, and forensics related to artificial intelligence. By discussing and addressing these issues proactively, we can work towards mitigating the risks associated with the malicious use of AI and safeguarding individuals and organizations from potential harm.

7. Conclusion

In conclusion, it is imperative to address the growing security threats and crimes associated with artificial intelligence through research and collaboration across various disciplines. The paper emphasizes the need for AI forensics to investigate and prevent malicious activities enabled by AI technology. By categorizing AI crime into AI as a tool crime and AI as a target crime, the taxonomy proposed in the document provides a framework for understanding and combating these evolving threats. Looking ahead, forensic stakeholders must enhance their expertise in AI to effectively tackle the challenges posed by AI crime. Overall, it is essential to uphold principles such as error documentation, transparency, reproducibility, and experience in AI forensics to ensure the reliability and accuracy of investigations. Furthermore, future research directions in AI forensics, including AI exploration, similarity analysis, adversarial attack detection, and damage assessment, must be explored to effectively combat AI-related crimes. By addressing these issues and fostering collaboration between forensic experts and AI professionals, we can work towards a safer and more secure AI-powered future.

References

[1.] S. V. Albrecht and P. Stone, ‘‘Autonomous agents modelling other agents: A comprehensive survey and open problems,’’ Artif. Intell., vol. 258, pp. 66–95, May 2018.

[2.] V. C. Müller and N. Bostrom, ‘‘Future progress in artificial intelligence: A survey of expert opinion,’’ in Fundamental Issues of Artificial Intelligence. Cham, Switzerland: Springer, 2016, pp. 555–572

[3] T. King, N. Aggarwal, M. Taddeo, and L. Floridi, ‘‘Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions,’’ SSRN Electron. J., vol. 26, no. 1, pp. 1–32, 2019.

[4] M. Brundage et al., ‘‘Malicious use of artificial intelligence: Forecasting, prevention, and mitigation,’’ 2018, arXiv: 18 02. 072 28. [Online]. Available: http://arxiv.org/abs/1802.07228

[5] N. Akhtar and A. Mian, ‘‘Threat of adversarial attacks on deep learning in computer vision: A survey,’’ IEEE Access, vol. 6, pp. 14410–14430, 2018.

[6] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain, ‘‘The science of fake news,’’ Science, vol. 359, no. 6380, pp. 1094–1096, 2018.

[7] H. Allcott and M. Gentzkow, ‘‘Social media and fake news in the 2016 election,’’ J. Econ. Perspect., vol. 31, no. 2, pp. 211–236, 2017.

[8] D. K. Citron and R. Chesney, Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy? Washington, DC, USA: Lawfare, 2018.

[9] H. Xue, S. Sun, G. Venkataramani, and T. Lan, ‘‘Machine learning-based analysis of program binaries: A comprehensive study,’’ IEEE Access, vol. 7, pp. 65889–65912, 2019.

[10] P. Mohassel and Y. Zhang, ‘‘SecureML: A system for scalable privacypreserving machine learning,’’ in Proc. IEEE Symp. Secur. Privacy (SP), May 2017, pp. 19–38.

[11] Y. Mao, S. Yi, Q. Li, J. Feng, F. Xu, and S. Zhong, ‘‘A privacy-preserving deep learning approach for face recognition with edge computing,’’ in Proc. USENIX Workshop Hot Topics Edge Comput. (HotEdge), 2018, pp. 1–6.

[12] A. Ilachinski, AI, Robots, and Swarms: Issues, Questions, and Recommended Studies. Arlington County, VA, USA: CNA Corporation, 2017.

[13] Q. Xiao, K. Li, D. Zhang, and W. Xu, ‘‘Security risks in deep learning implementations,’ ’ in Proc. IEEE Secur. Privacy Workshops (SPW), May 2018, pp. 123–128.

[14] Corona, G. Giacinto, and F. Roli, ‘‘Adversarial attacks against intrusion detection systems: Taxonomy, solutions and open issues,’’ Inf. Sci., vol. 239, pp. 201–225, Aug. 2013.

[15] A. Nguyen, J. Yosinski, and J. Clune, ‘‘Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 427–436.

[16] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, ‘‘Robust physical-world attacks on deep learning visual classification,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1625–1634.