International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

ARTIFICIAL INTELLIGENCE CRIME AN OVERVIEW OF MALICIOUS USE AND ABUSE OF AI

Dasari Narasimha Das[1], V.Abdul Razzaq[2]

[1]M. Tech Student, Department of Computer science and engineering, Golden valley integrated campus, Kadiri Road, Angallu Post, Madanapalli, Chittoor, Andhra Pradesh 517326
[2]Assistant Professor, Department of Computer science and engineering, Golden valley integrated campus, Kadiri Road, Angallu Post, Madanapalli, Chittoor, Andhra Pradesh 517326

Abstract: Artificial Intelligence (AI) has rapidly transformed various sectors by providing intelligent solutions for automation, decision-making, prediction, and information processing. However, the same capabilities can also be misused for criminal and malicious activities. **Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI** examines the emerging risks associated with the intentional misuse of AI technologies for cybercrime, fraud, identity theft, misinformation, social engineering, automated attacks, and other unlawful activities. AI can enable attackers to generate convincing fraudulent content, automate large-scale scams, manipulate digital media, and identify vulnerabilities more efficiently than traditional methods.

This study provides an overview of the major forms of AI-enabled crime and analyzes how technologies such as machine learning, generative AI, deepfakes, natural language processing, and automated decision systems can be exploited. It also discusses the challenges faced by individuals, organizations, law-enforcement agencies, and policymakers in identifying and responding to AI-assisted crimes. In addition, the study highlights the importance of AI-based detection mechanisms, cybersecurity practices, digital forensics, ethical AI development, user awareness, and appropriate regulatory frameworks. The objective is to improve understanding of the criminal misuse of AI and encourage the development of responsible, secure, and trustworthy AI systems that minimize potential harm while preserving the legitimate benefits of artificial intelligence.

1. INTRODUCTION

Artificial Intelligence (AI) has become one of the most influential technologies of the modern digital era, transforming areas such as healthcare, education, finance, transportation, communication, and cybersecurity. AI systems can analyze large amounts of data, recognize patterns, generate content, automate tasks, and make decisions with limited human intervention. Although these capabilities provide significant benefits to society, they can also be exploited by individuals and criminal organizations for malicious purposes. The misuse of AI has created a new category of threats commonly referred to as **AI-enabled or AI-assisted crime**.

The rapid development of generative AI, deep learning, natural language processing, computer vision, and autonomous systems has increased the possibilities for criminal exploitation. Attackers can use AI to create realistic fake images, videos, audio, and text, automate phishing and social engineering attacks, generate malicious content, impersonate individuals, manipulate information, and conduct fraud on a larger scale. Deepfake technology, for example, can be misused to create convincing digital identities or fabricated events, making it increasingly difficult for people and organizations to distinguish genuine information from manipulated content.

AI can also increase the speed, scale, and sophistication of traditional cybercrime. Criminals may use automated systems to identify potential targets, analyze publicly available information, personalize fraudulent messages, and adapt their strategies based on victims' responses. This creates challenges for conventional security mechanisms because AI-assisted attacks can be more dynamic and difficult to detect. At the same time, AI itself can become a target of attacks through techniques such as data poisoning, adversarial manipulation, model theft, and unauthorized access.

The impact of AI-related crime extends beyond financial losses and cybersecurity incidents. It can affect personal privacy, reputation, public trust, organizational security, and the reliability of digital information. The increasing availability of powerful AI tools makes it important for governments, technology companies, educational institutions, cybersecurity professionals, and users to understand the potential consequences of their misuse.

Therefore, studying the **malicious use and abuse of Artificial Intelligence** is essential for developing effective prevention and detection strategies. This topic provides an overview of major AI-related criminal activities, their potential impacts, existing challenges, and possible approaches for mitigation. Responsible AI development, cybersecurity awareness, digital forensics, robust authentication, content verification, ethical guidelines, and appropriate legal frameworks can play an important role in reducing AI-enabled crime while ensuring that AI technologies continue to be used for beneficial and legitimate purposes.

2. EXISTING SYSTEM

The existing system for addressing **AI-enabled crime and the malicious use of Artificial Intelligence** mainly depends on conventional cybersecurity mechanisms, manual investigation, rule-based detection systems, and platform-level content moderation. Organizations generally use firewalls, antivirus software, intrusion detection systems, spam filters, authentication mechanisms, and fraud detection tools to identify suspicious activities. Law-enforcement agencies and cybersecurity teams also rely on digital forensics and manual analysis to investigate cybercrime incidents. Social media and online platforms use automated moderation and content verification techniques to detect fake or harmful content. However, these approaches were primarily designed for traditional cyber threats and are not always capable of effectively handling rapidly evolving AI-assisted attacks.

2.1 Disadvantages of Existing System

- **Difficulty detecting AI-generated content:** Traditional detection methods may struggle to accurately identify sophisticated AI-generated text, images, audio, and videos.
- **Limited adaptability:** Rule-based security systems depend on predefined patterns and signatures, making them less effective against new and constantly changing AI-assisted attacks.
- **High false-positive and false-negative rates:** Automated detection systems may incorrectly classify legitimate activities as malicious or fail to detect sophisticated attacks.
- **Increasing attack automation:** AI allows attackers to automate phishing, fraud, impersonation, and information manipulation at a much larger scale than traditional methods.
- **Dependence on manual investigation:** Complex AI-related crimes often require expert analysis, digital forensics, and significant human effort, which can increase investigation time.
- **Difficulty identifying deepfakes:** Highly realistic synthetic media can make it difficult for users and investigators to distinguish genuine content from manipulated content.
- **Privacy concerns:** Monitoring and analyzing large amounts of user and network data can create privacy and data-protection challenges.
- **Rapid evolution of AI threats:** Security mechanisms may become outdated quickly

because attackers continuously develop new methods for abusing emerging AI technologies.

- **Lack of standardized regulations:** Differences in laws, policies, and technical standards across regions can make the investigation and prevention of AI-enabled crimes more complicated.
- **Limited public awareness:** Many users are not sufficiently aware of AI-generated scams, impersonation techniques, deepfakes, and other emerging threats, increasing their vulnerability.

3. PROPOSED SYSTEM

The proposed system aims to develop an intelligent framework for identifying, analyzing, and mitigating the malicious use and abuse of Artificial Intelligence. The system combines **Artificial Intelligence, Machine Learning, Natural Language Processing, Deep Learning, and digital forensic techniques** to analyze suspicious activities and identify potential AI-enabled crimes. The system can collect relevant data from sources such as messages, emails, online content, digital media, and security logs, followed by preprocessing and feature extraction to identify suspicious patterns.

The proposed framework uses machine learning and deep learning models to detect activities associated with phishing, online fraud, impersonation, misinformation, malicious content, and AI-generated media. For text-based analysis, Natural Language Processing techniques can identify suspicious language patterns, fraudulent messages, and potentially manipulated information. For images, audio, and video, deep learning-based detection techniques can analyze digital characteristics and identify possible synthetic or manipulated content. The system can also maintain logs of detected incidents to support further investigation and digital forensic analysis.

A risk-analysis component can classify detected activities according to their potential threat level and generate alerts for suspicious behaviour. The system can continuously learn from newly identified patterns, allowing it to adapt to emerging forms of AI-assisted crime. An administrator or security analyst can monitor detected incidents through a centralized dashboard and review the evidence generated by the system.

3.1 Advantages of Proposed System

- **Intelligent threat detection:** Uses AI and machine learning to identify suspicious patterns more effectively.
- **Detection of AI-generated content:** Supports analysis of potentially manipulated text, images, audio, and video.
- **Early identification of threats:** Generates alerts when suspicious activities are detected.
- **Automated analysis:** Reduces the amount of manual effort required for preliminary investigation.
- **Adaptive security:** Machine learning models can be updated with new threat patterns.
- **Multi-source analysis:** Can analyze different types of digital information and security logs.
- **Improved investigation:** Maintains detected incidents and relevant analysis results for forensic examination.
- **Reduced response time:** Automated detection and alert generation can help security teams respond more quickly.
- **Scalable framework:** The system can be extended to accommodate new AI-related crime patterns and detection techniques.
- **Enhanced awareness:** Provides useful information about emerging AI threats and supports safer use of AI technologies.

4. SYSTEM ANALYSIS

The proposed **Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI** system is analyzed to determine how effectively AI-based techniques can be used to identify, classify, and monitor different forms of AI-enabled criminal activities. The system focuses on collecting relevant digital data, preprocessing the collected information, extracting important features, and applying machine

learning or deep learning techniques to identify suspicious patterns. The analysis considers the ability of the system to handle different forms of malicious activities, including phishing, online fraud, impersonation, misinformation, deepfake content, and other AI-assisted cyber threats.

The system first receives data from different digital sources such as emails, messages, websites, uploaded media, and security logs. The collected data is cleaned and transformed into a suitable format for analysis. Feature extraction techniques are then applied to identify characteristics that can distinguish normal activities from suspicious or malicious activities. For text-based information, Natural Language Processing techniques can be used to analyze keywords, sentence patterns, and contextual information. For multimedia content, suitable deep learning approaches can examine visual, audio, or other digital features associated with manipulated content.

The processed data is provided to trained machine learning or deep learning models for classification and prediction. The model categorizes the analyzed information as legitimate, suspicious, or potentially malicious based on learned patterns. When a high-risk activity is identified, the system generates an alert and stores the incident details for further investigation. A monitoring interface can provide administrators or security analysts with information about detected threats, their categories, severity levels, and analysis results.

The system analysis also considers **performance, accuracy, scalability, security, and usability**. Since AI-enabled crimes continuously evolve, the proposed system should support regular model updates and incorporation of newly identified threat patterns. It should also minimize false detections while maintaining strong detection capability. Proper access control, secure storage, privacy protection, and responsible handling of analyzed data are important components of the system.

Overall, the system analysis demonstrates that an AI-assisted crime detection framework can provide a more adaptive approach than conventional rule-based security mechanisms. By combining data analysis, machine learning, deep learning, NLP, and automated alert mechanisms, the proposed system can support early identification of suspicious activities and assist cybersecurity professionals in investigating emerging forms of AI-related crime.

5. SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

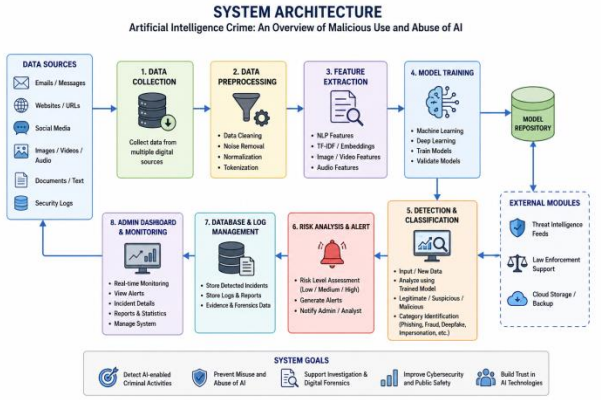
- **Processor:** Intel Core i5 or equivalent and above
- **RAM:** Minimum 8 GB, preferably 16 GB
- **Hard Disk:** Minimum 500 GB
- **GPU:** Optional NVIDIA GPU for deep learning and multimedia analysis
- **Display:** Standard HD monitor
- **Keyboard and Mouse:** Standard input devices
- **Internet Connection:** Required for dataset collection, software installation, and updates

Software Requirements

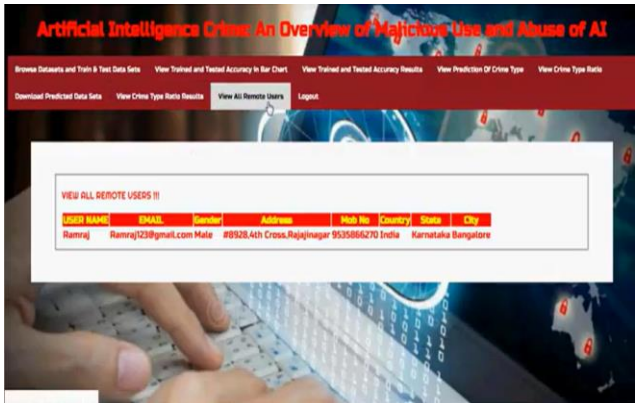
- **Operating System:** Windows 10/11, Linux, or Ubuntu
- **Programming Language:** Python 3.x
- **Development Environment:** Visual Studio Code / PyCharm / Jupyter Notebook
- **Database:** MySQL or SQLite
- **Web Framework:** Django or Flask, if a web-based interface is implemented
- **Machine Learning:** Scikit-learn
- **Deep Learning:** TensorFlow / Keras or PyTorch
- **Data Processing:** Pandas and NumPy
- **Data Visualization:** Matplotlib
- **Natural Language Processing:** NLTK / spaCy / Transformers
- **Image Processing:** OpenCV and Pillow

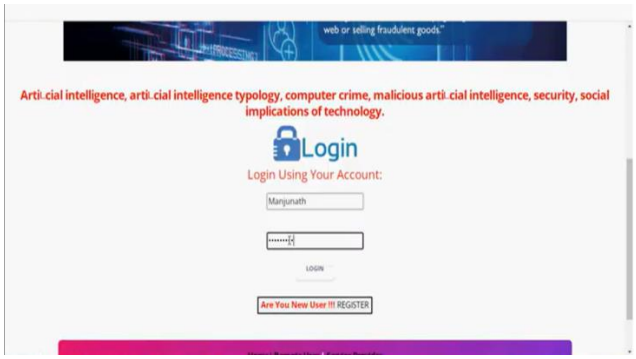
- **Model Storage:**Joblib or Pickle
- **Web Technologies:** HTML5, CSS3, JavaScript, and Bootstrap

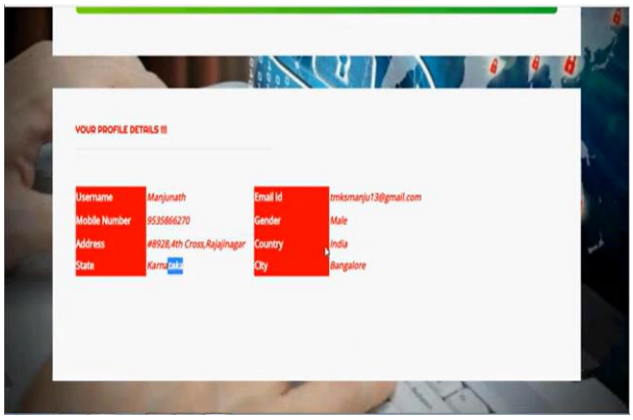
System Architecture:

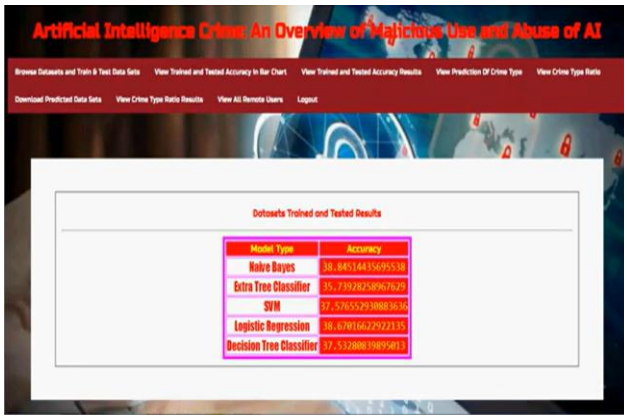


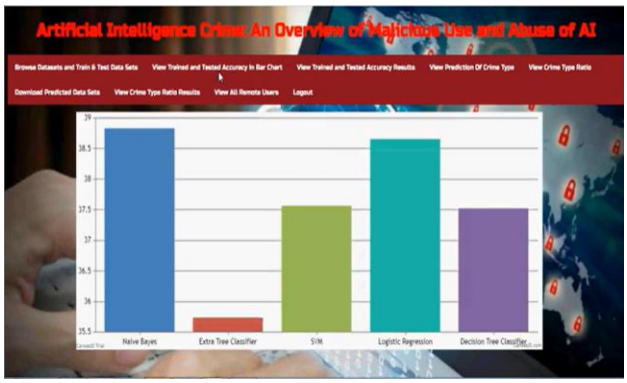
6. RESULTS AND ANALYSIS

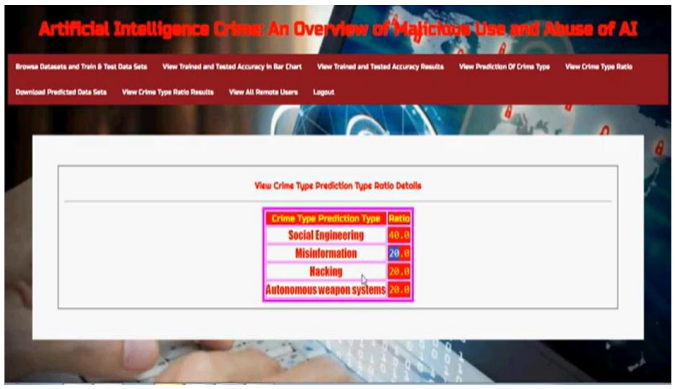












7. CONCLUSIONS

The **Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI** project highlights the growing risks associated with the misuse of Artificial Intelligence for criminal and malicious activities. Although AI provides significant benefits in automation, prediction, communication, and decision-making, its misuse can support phishing, fraud, impersonation, misinformation, deepfakes, and other forms of cybercrime. Traditional security mechanisms alone may not be sufficient to address these rapidly evolving threats.

The proposed system uses **machine learning, deep learning, Natural Language Processing, and digital data analysis** to identify suspicious activities and potentially malicious content. By collecting relevant data, preprocessing it, extracting meaningful features, and applying intelligent classification models, the system can assist in detecting and categorizing AI-enabled criminal activities. Automated alerts, risk analysis, and incident storage can further support security analysts in monitoring and investigating suspicious events.

Overall, the project demonstrates the importance of developing intelligent and adaptive security mechanisms to address the emerging challenges created by malicious AI usage. Regular model updates, reliable datasets, privacy protection, ethical AI practices, and effective cybersecurity policies are essential for improving the system. In the future, the proposed approach can be enhanced with real-time detection, multimodal deepfake analysis, explainable AI, advanced digital forensics, and continuously learning models to provide stronger protection against the evolving landscape of AI-enabled crime.

REFERENCES

1. Russell, S., & Norvig, P. *Artificial Intelligence: A Modern Approach*. Pearson Education.
2. Goodfellow, I., Bengio, Y., & Courville, A. *Deep Learning*. MIT Press.
3. Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer.
4. Russell, S. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.
5. Floridi, L., & Cowls, J. "A Unified Framework of Five Principles for AI in Society." *Harvard Data Science Review*, 2019.
6. Chesney, R., & Citron, D. K. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review*, 2019.
7. Mirsky, Y., & Lee, W. "The Creation and Detection of Deepfakes: A Survey." *ACM Computing Surveys*, 2021.
8. Europol. *Facing Reality? Law Enforcement and the Challenge of Deepfakes and Generative AI*. European Union Agency for Law Enforcement Cooperation.
9. National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce, 2023.
10. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational, Scientific and Cultural Organization, 2021.
11. Taddeo, M., McCutcheon, T., &Floridi, L. "Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword." *Nature Machine Intelligence*, 2019.
12. Kshetri, N. "The Evolution of Cybercrime: From Cybercrime to AI-Enabled Cybercrime." Relevant cybersecurity and AI-crime research literature.
13. Sarker, I. H. "Machine Learning: Algorithms, Real-World Applications and Research Directions." *SN Computer Science*, 2021.
14. Akhtar, Z., & Dasgupta, D. Research literature on adversarial machine learning, AI security, and malicious use of artificial intelligence.
15. Brundage, M., et al. "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation." *arXiv preprint arXiv:1802.07228*, 2018.