Artificial Intelligence Crime: An Overview Of Malicious Use And Abuse Of AI

Veerender Aerranagula[1] R.satish[2], A.yogi manikanta[3], K.deveena[4]

[1]Assistant professor, Dept of CSE(Data science)

[2, 3, 4]Dept of CSE(Data science)

[1, 2, 3, 4] CMR Technical Campus, Kandlakoya, Hyderabad, Telangana, India.

Abstract- *The rapid evolution of Artificial Intelligence (AI) has introduced new threats and expanded existing vulnerabilities. This article reviews the malicious use and abuse of AI, constructing a typology of threats. Four types of malicious abuse of AI are identified: integrity attacks, unintended AI outcomes, algorithmic trading, and membership inference attacks. Additionally, four types of malicious use of AI are highlighted, social engineering, misinformation, hacking, and autonomous weapon systems. Enhanced collaboration is vital to minimize risks and avoid harmful consequences.*

Keywords- Crime AI, Cybersecurity, Machine Learning, Malicious Use, Detection

I. INTRODUCTION

The rapid evolution of Artificial Intelligence (AI) has led to its widespread adoption across various sectors, but it has also created new avenues for malicious use and abuse. The project "Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI" aims to explore the darker side of AI, where it is exploited for criminal activities, expanding existing vulnerabilities and introducing new threats. The malicious abuse of AI can take various forms, including integrity attacks, unintended AI outcomes, algorithmic trading, and membership inference attacks. Furthermore, AI can be used for malicious purposes such as social engineering, spreading misinformation or fake news, hacking, and developing autonomous weapon systems. The security risks associated with AI systems are also a major concern, as they can inherit existing computer system security vulnerabilities and be susceptible to unique cyber attacks aided by AI. The project's primary objective is to identify the vulnerabilities of AI models and outline how malicious actors can abuse them, as well as explore AI-enabled and AI-enhanced attacks. By developing a typology of malicious use and abuse of AI systems, the project seeks to clarify the types of activities and corresponding risks involved. The research outcomes are expected to provide a comprehensive overview of AI-related crimes, enabling advanced reflection on governance strategies, policies, and activities to minimize risks and avoid harmful consequences. The project's findings will contribute to a better understanding of the malicious use and abuse of AI, ultimately informing the development of effective countermeasures to mitigate these threats. As AI continues to evolve and become increasingly integrated into various aspects of life, it is essential to address the risks associated with its malicious use and abuse, ensuring that its benefits are realized while minimizing its negative consequences. By exploring the intersection of AI and crime, this project aims to provide valuable insights into the complex and rapidly evolving landscape of AI-related threats, ultimately contributing to the development of a safer and more secure digital environment. The project's outcomes will have significant implications for policymakers, industry leaders, and researchers, providing them with a comprehensive understanding of the malicious use and abuse of AI and informing the development of effective strategies to counter these threats. Ultimately, the project's findings will help to ensure that the benefits of AI are realized while minimizing its risks, promoting a safer and more secure digital environment for all.

II. PROCEDURE FOR PAPER SUBMISSION

A. Integrity Attacks

Adversarial attacks manipulating AI inputs / outputs (e.g., deep fake fraud).

B. Unintended Outcomes

Bias amplification or flawed decision-making in high-stakes domains (e.g., healthcare).

III. MALICIOUS USE OF AI

A. Social Engineering AI-Powered Phishing

AI-powered phishing scams utilize advanced technology to deceive individuals into revealing sensitive information. Voice cloning, a form of deep fake technology, is a significant threat in this domain. Scammers can replicate a person's voice to gain trust and extract confidential data. This technique is particularly effective in phishing attacks, as it exploits the natural human tendency to trust familiar voices.

There are various types of AI-powered phishing attacks, including voice cloning scams, deep fake emails and messages, and AI-generated spam. These attacks can be challenging to detect, making them a significant concern for individuals and organizations alike.

B. Autonomous Weapons: Ethical and Security Dilemmas

Autonomous weapons, powered by AI, raise significant ethical and security concerns. These systems can make life-or-death decisions without human intervention, sparking debates about accountability and the potential for unintended consequences. The lack of accountability, unintended consequences, and escalation of conflicts are key concerns surrounding autonomous weapons. The development and deployment of these systems require careful consideration of these ethical and security dilemmas to ensure that their use aligns with international humanitarian law and human values. As autonomous weapons continue to evolve, it is essential to address these concerns and establish clear guidelines for their development and use.

IV. MITIGATION STRATEGIE

Mitigation strategies against AI misuse require a dual approach combining policy and technology. Regulatory frameworks like the **EU AI Act** ban harmful applications while mandating transparency for high-risk systems. Technical solutions include **adversarial training** to counter data manipulation, **explainable AI (XAI)** for accountability, and **federated learning** to preserve privacy. Cross-sector collaboration through initiatives like the **Partnership on AI** and **NIST's Risk Management Framework** further strengthens defenses by promoting ethical standards and threat-sharing among governments, researchers, and corporations.

V. FIGURES

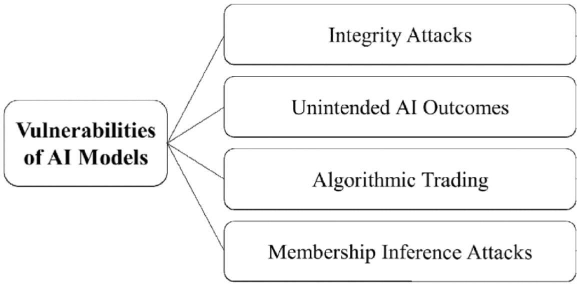

Fig .1 Adversarial Manipulation of AI systems

The graphic highlights four major AI risks: (1) Integrity Attacks involving adversarial manipulation of AI systems, (2) Unintended Outcomes like biased or erroneous decisions, (3) inherent Vulnerabilities enabling data breaches or model hijacking, and (4)Algorithmic Trading exploits that manipulate financial markets through AI-driven strategies. Together, these threats underscore the need for robust safeguards.

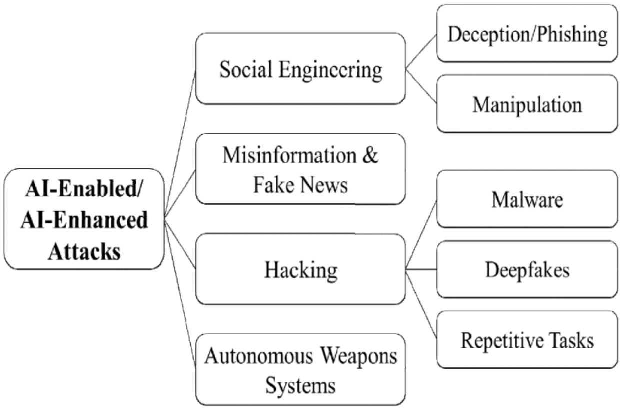

Fig 2. Autonomous Weapons System

May span both columns. Place figure captions below the figures; place of titles above the figure has two parts, include the labels (a) and The graphic outlines AI-powered social engineering threats, including AI-enhanced phishing/deception, deep fake-driven misinformation, autonomous weapons systems, and automated hacking/malware attacks. These tools enable hyper-personalized manipulation, mass disinformation campaigns, and scalable cybercrime, demonstrating how AI amplifies traditional threats while creating dangerous new attack vectors requiring urgent countermeasures.

VI. CONCLUSION

The threats posed by the use and abuse of AI systems must be well understood to create mechanisms that protect society and critical infrastructures from attacks. Based on the available literature, reports, and previous incidents, we focused on creating a classification of how AI systems can be used or abused by malicious actors. This includes, but is not limited to, physical, psychological, political, and economic harm. We explored the vulnerabilities of AI models, such as unintended outcomes, and AI-enabled and AI-enhanced attacks, such as forgery. This article also describes past incidents, such as the 2010 flash crash and the Cambridge Analytical scandal, manifesting the challenges at hand. We also outlined attacks that, to the best of our knowledge, have only been demonstrated through "proof of concept", such as IBM's Deep Locker. In response to the risks presented in this paper, we have also explored some possible mitigation strategies. Industries, governments, civil society, and individuals should cooperate in developing knowledge and raising awareness while developing technical and operational systems and procedures to address the challenges. Ernounsand element symbols.

REFERENCES

[1] Brundidge, M., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.arXiv:1802.07228.(Seminal report on AI threats)

[2] European Commission. (2021). Proposal for a Regulation on Artificial Intelligence (EU AI Act). Brussels.(Key regulatory framework)

[3] C.Y.LGoodfellow,I.etal. (2015). Explaining and Harnessing Adversarial Examples. ICLR.(Foundational paper on adversarial attacks)

[4] O'Neil,C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality.Crown.(Bias/unintendedconsequences)

[5] Taddeo, M., &Floridi, L. (2018). Regulate Artificial Intelligence to Avert Cyber Arms Race.Nature,556(7701),296-298.(Autonomous weapons)

[6] NIST. (2023). AIRiskManagementFramework (AI RMF 1.0). NIST.(Technical standards)

[7] Hao, K. (2021). How AI-Powered Deepfakes Could Upend Elections. MIT Tech Review.(Misinformation case study)

[8] Barreno, M., et al. (2010). Can Machine Learning Be Secure? ACM ASIACCS.(Early vulnerabilities analysis)

[9] S.PartnershiponAI. (2022). Recommendations for AI and Social Engineering. PAIR.(Industry guidelines)

[10] Schneier, B. (2020). Click Here to Kill Everybody: Security in an AI-Enabled World. Norton.(Policy/ethics perspective)

[11] Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New FrontierofPower.PublicAffairs.(Social impacts of AI/data exploitation)

[12] Marcus, G. & Davis, E. (2019). Rebooting AI: Building Artificial Intelligence We Can Trust. Vintage.(AI safety and reliability challenges)

[13] Cath,C.(2018).GoverningArtificialIntelligence: Ethical, Legal and Technical Opportunities and Challenges. Philosophical Transactions of the Royal Society A.(Regulatory approaches to AI)

[14] Susskind, J. (2018). Future Politics: Living Together in a World Transformed by Tech. Oxford University Press.(Political implications of AI/automation)

[15] Osoba, O.A. & Welser IV, W. (2017). The Risks of Artificial Intelligence to Security and the Future of Work. RAND Corporation.(Economic and security risk analysis)