Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 9, September : 2024

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

[1]Kaskurthy laxmi devi
Mtech.Department of computer science and engineering,
Vaagdevi college of engineering, Bollikunta,Warnagal,Telnagana.
Email: laxmidevikaskurthy@gmail.com

[2]Dr.p.shailaja
Assoc. Prof. Department of computer science and engineering,
Vaagdevi college of engineering, Bollikunta,Warnagal,Telanagana.
Email: Pokalashylaja@gmail.com

[3]Dr. E. Balakrishna
Assoc. Prof. Department of computer science and engineering,
Vaagdevi college of engineering, Bollikunta,Warnagal,Telanagana.
Email: balakrishnakits@gmail.com

[4]Dr. N. Satyavathi
Head of Department of computer science and engineering,
Vaagdevi college of engineering, Bollikunta,Warnagal,Telanagana.
Email: satyanadendla15@gmail.com

ABSTRACT The capabilities of Artificial Intelligence (AI) evolve rapidly and affect almost all sectors of society. AI has been increasingly integrated into criminal and harmful activities, expanding existing vulnerabilities, and introducing new threats. This article reviews the relevant literature, reports, and representative incidents which allows to construct a typology of the malicious use and abuse of systems with AI capabilities. The main objective is to clarify the types of activities and corresponding risks. Our starting point is to identify the vulnerabilities of AI models and outline how malicious actors can abuse them. Subsequently, we explore AI-enabled and AI-enhanced attacks. While we present a comprehensive overview, we do not aim for a conclusive and exhaustive classification. Rather, we provide an overview of the risks of enhanced AI application, that contributes to the growing body of knowledge on the issue. Specifically, we suggest four types of malicious abuse of AI (integrity attacks, unintended AI outcomes, algorithmic trading, membership inference attacks) and four types of malicious use of AI (social engineering, misinformation/fake news, hacking, autonomous weapon systems). Mapping these threats enables advanced reflection of governance strategies, policies, and activities that can be developed or improved to minimize risks and avoid harmful consequences. Enhanced collaboration among governments, industries, and civil society actors is vital to increase preparedness and resilience against malicious use and abuse of AI.

INDEX TERMS Artificial intelligence, artificial intelligence typology, computer crime, malicious artificial intelligence, security, social implications of technology

I. INTRODUCTION

The impact of systems using Artificial Intelligence (AI) is at the center of numerous academic studies, political debates, and reports of civil society organizations. The development of AI has become the subject of praise due to unprecedented technological capabilities, such as enhanced possibilities for automated image recognition (e.g., detection of cancer in the field of medicine). However, it has also been criticized— even feared—due to aspects such as the uncertain consequences of automation for the labor market (e.g., concerns of mass unemployment). This duality of positive vs. negative aspects of the technology can also be identified in the context of cybersecurity and cybercrime. Governments use AI to enhance their capabilities, whereas the same technology can be used for attacks against them. While the recent surge in AI development has been fueled by the private sector and applications in customer-oriented applications, sectors such as defense might use similar capabilities in their operations. At the same time, it is increasingly difficult to distinguish between the actions of state and non-state actors. This has recently been demonstrated by a wave of ransomware attacks targeting public infrastructure in many countries, such as the Colonial Pipeline in the United States in May 2021. Additionally, programs and applications developed for non-malicious purposes can also be implemented or modified for malicious intent and potentially cause harm. The dual-use aspect of technology is not an entirely new problem when it comes to cybercrime or (cyber-)security. Nevertheless, how AI can be leveraged for malicious use and abuse constitutes novel vulnerabilities.

A. BACKGROUND

The integration of AI in various sectors has led to both commendable advancements and significant concerns. On one hand, AI technologies like automated image recognition have revolutionized fields such as medicine by improving diagnostic accuracy and efficiency. On the other hand, the automation potential of AI poses a threat to the labor market, raising fears of mass unemployment. This dichotomy extends into the realm of cybersecurity and cybercrime. Governments and institutions harness AI to bolster their security measures, yet the same technologies can be exploited for cyberattacks against them. The private sector has primarily driven AI advancements, particularly for customer-focused applications, but defense sectors also utilize similar technologies. The challenge of distinguishing between state and non-state actors in cyber activities further complicates the scenario, as evidenced by ransomware attacks like the Colonial Pipeline incident in May 2021. Furthermore, AI applications intended for benign purposes can be repurposed for malicious activities, introducing new vulnerabilities.

B. MOTIVATION

This project is motivated by the need to understand and address the dual-use nature of AI technologies in cybersecurity and cybercrime. As AI continues to evolve, its potential for both beneficial and harmful applications becomes increasingly evident. The goal is to develop a comprehensive understanding of how AI can be misused in cyber contexts and to create a typology that categorizes harmful AI-based activities. By doing so, we aim to equip cybersecurity organizations and governmental agencies with the knowledge needed to anticipate and prepare for malicious AI-driven incidents. Enhancing preparedness against such threats is crucial for maintaining security and resilience in an increasingly digital world.

C. OBJECTIVES

The primary objective of this project is to evaluate the main categories of use and abuse of AI in a criminal context and develop a typology that catalogs harmful AI-based activities.

By achieving these objectives, this project aims to contribute to the broader effort of safeguarding digital infrastructures against the growing threat of AI-enhanced cybercrime.

II. OVERVIEW OF MALICIOUS USE AND ABUSE OF AI

After analyzing the academic literature (43 papers, books, and conference proceedings), reports (5 reports), and other documents (26 sources, including news stories, web pages, and other general documents), it was possible to identify the main malicious uses and abuses of AI systems.

MALICIOUS ABUSE OF AI: VULNERABILITIES OF AI MODELS

1) INTEGRITY ATTACKS

Machine learning (ML) has become more prevalent in recent years. This has created incentives for attackers to manipulate models (e.g., the software itself) or the underlying data, making ML models prone to integrity attacks. In integrity attacks, hackers attempt to inject false information into a system to corrupt the data, undermining their trustworthiness. One of the risks associated with the vulnerability of AI models is the creation of ‘adversarial examples’. According to, ‘‘adversarial examples are malicious inputs designed to fool machine learning models’’ which causes misclassification of material scrutinized by the systems. In some cases, the perturbations are too subtle to be perceived by human observers, but they still cause AI systems to make mistakes. One example of an adversarial ML is a ‘poisoning attack’. The attacker influences the training data of the system to alter the results of a predictive model by injecting a few corrupted points in the training process. In other words, poisonous samples can be injected into the training data to manipulate the classifier, leading to undesirable consequences. A concrete example is the attack on Tay, Microsoft’s AI chatbot, which was released in 2016. The chatbot had the objective of creating tweets that could not be distinguished from a human actor. Within a few hours of release, users launched a coordinated attack in which they tweeted offensive words and phrases, exploring Tay’s “repeat after me” function. This led the bot to reproduce similarly objectionable content . According to [36], the Corporate VicePresident of Microsoft, “although we had prepared for many types of abuses of the system, we had made a critical oversight for this specific attack.” Consequently, after less than 16 hours, Microsoft had to suspend the account. This demonstrates that defending a chatbot against attacks is challenging, especially when the system is trained in online environments with unforeseeable live interactions. Researchers at New York University (NYU) explored another risk associated with the context of outsourced training data. They demonstrated that an adversary might create a BadNet (a maliciously trained network), which displays conventional behavior until a potential attacker triggers an attack. To test this hypothesis, BadNets were implemented in a complex traffic sign detection system. They demonstrated that a stop sign could be correctly identified by a selfdriving car until a stop sign with a pre-defined trigger (yellow ‘Post-It’ note) was presented. This study demonstrates that AI models might be susceptible to data poisoning and adversarial examples, resulting in misclassifications and errors with potentially grave consequences that are difficult to foresee for humans unfamiliar with the technology. This might be one of the reasons why the recently proposed EU AI Act entails specific requirements for training data of ‘high-risk systems’ in Article 10

2) UNINTENDED OUTCOMES OF THE USE OF AI

Models used to train AI systems can present a different result from what was expected by the developer for various reasons. For instance, models based on neural networks may unintentionally memorize and disclose details. This can be problematic, especially when the data used to train the models are private or sensitive.explained the phenomenon: during the learning process, such models might memorize details unrelated to the primary task. To prevent harmful consequences from unintended memorization and disclosure of information by the algorithm, it is necessary to apply techniques that guarantee data privacy. The team behind the development of Smart Compose, the real-time suggestion system used by Google’s Gmail service, considered this carefully. To avoid unintended memorization, they conducted “extensive testing to make sure that only common phrases used by multiple users are memorized”. Their goal was to prevent the models from learning details (e.g., private information) that were not related to the primary task (e.g. general and commonly used phrases) while training the algorithm. For example, when a user enters a text prefix such as “my ID number is”, the model should not suggest a text completion with the ID number of another user . This challenge serves as one example in which the developer does not have the malicious intent of disclosing the user’s personal information; the potential harm resides in the possibility that the model performs differently than previously expected (i.e., by memorizing private data).

III.LITERATURE SURVEY

In the ever-evolving landscape of artificial intelligence (AI) research, a series of seminal papers have emerged, each shedding light on the potential risks and challenges associated with the malicious use of AI. One such landmark paper, titled "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," authored by Brundage et al., stands out as a comprehensive survey of the security threats posed by AI and proposes strategies to forecast, prevent, and mitigate these threats. Published in [Year], this paper delves into the ways in which AI may impact various security domains, including digital security, physical security, and political security, emphasizing the urgent need for further research and collaboration to address emerging challenges.

In a parallel vein, Mittelstadt et al.'s paper, "Ethical and Social Implications of Artificial Intelligence for Crime, Policing, and Courts," published [Year], offers a critical examination of the ethical and social implications of AI technologies in crime prevention, policing, and the criminal justice system. Drawing on insights from ethical theory, legal analysis, and empirical research, the authors delve into issues such as bias and discrimination in AI algorithms, the impact of AI on privacy and civil liberties, and the challenges of ensuring accountability and transparency in AI-enabled law enforcement practices.

Complementing these efforts, Cave et al.'s paper, "The Dark Side of Artificial Intelligence: A Framework for Understanding and Addressing Malicious Use," [Year], presents a comprehensive framework for understanding and addressing the malicious use of AI. Through an interdisciplinary lens encompassing computer science, cognitive psychology, and political science, the authors analyze various ways in which AI technologies can be exploited for nefarious purposes, from cybercrime to disinformation campaigns and autonomous weapons systems. Their proposed multi-dimensional approach advocates for technical safeguards, regulatory frameworks, and international cooperation to mitigate these risks effectively.

Meanwhile, Panchenko et al.'s paper, "AI-Enabled Cyber Attacks: Assessing the Security Risks and Countermeasures," [Year], delves into the specific realm of cybersecurity, investigating the security risks posed by AI-enabled cyber attacks and proposing countermeasures to mitigate these risks. Through empirical analysis and theoretical modeling, the authors identify vulnerabilities in existing cybersecurity systems and advocate for defensive strategies such as anomaly detection algorithms and adversarial training techniques.

Finally, Goodfellow et al.'s paper, "Adversarial Attacks and Defenses in Deep Learning," [Year], focuses on the vulnerabilities inherent in deep learning systems and explores techniques for generating adversarial examples that can fool neural networks. The authors discuss strategies for mitigating these attacks through robust training methods and propose avenues for future research to enhance AI security.

Together, these seminal papers represent a collective effort to understand, anticipate, and mitigate the malicious use of AI, shaping the discourse and guiding efforts to promote the responsible development and deployment of AI technologies.

IV. EXISTING SYSTEM

To build on previous work and expand the understanding of how AI broadens the potential for malicious activities online, this article evaluates the main categories of use and abuse of AI in a criminal context. We provide several salient examples that allow us to illustrate the challenges at hand. Based on these examples, we present a typology that catalogs the main harmful AI-based activities. Developing knowledge and understanding about the potential malicious use and abuse of AI enables cybersecurity organizations and governmental agencies to anticipate such incidents and increase their preparedness against attacks. Furthermore, a typology is greatly useful in structuring research efforts and identifying gaps in knowledge in areas where more research is warranted.

V. PROPOSED SYSTEM

This paper explores the malicious use and abuse of AI systems, providing a typology of concepts, threat scenarios, and possibilities. It aims to develop preventive measures and proactive responses, fostering interdisciplinary collaboration between STEM fields, legal practitioners, and policymakers. The study uses extensive analysis of cybercrime literature and identifies various malicious activities, particularly the manipulation of machine learning models and data. Integrity attacks are a significant concern.

VI.SYSTEM DESIGN

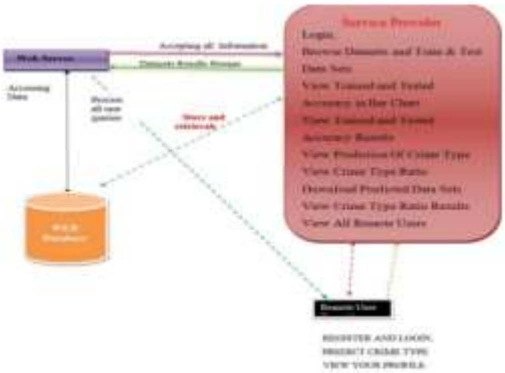

Fig1: Architecture of system.

A. MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View, Prediction Of Crime Type, View Crime Type Ratio, Download Predicted Data Sets, View Crime Type Ratio Results, View All Remote Users. In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT CRIME TYPE, VIEW YOUR PROFILE.

B. MAIN OBJECTIVES

To establish adequate security measures against AI-related attacks, it is necessary to comprehend the different types of malicious use and abuse of AI and map it, including the corresponding risks. However, there is a general lack of comprehensive and interdisciplinary assessment of the types of AI-enabled and AI-dependent cyberattacks, which might negatively affect the development of measures against them. Consequently, data security, personal safety, and political stability are at stake. This is to classify different types of malicious AI to expand the body of knowledge on the subject in a more holistic manner.

Specifically, this research aims to propose a typology of the malicious use and abuse of AI based on empirical evidence and contemporary discourse, analyzing how AI systems are used to compromise confidentiality, integrity, and data availability. The technique of classification of similar subjects into groups has been established for more than 2000 years,and such a study can be the starting point for the development of granular and in-depth analysis. Thus, our objectives are limited to identifying essential elements of the malicious use and abuse of AI, and to collect evidence of their use in practice. The compiled data enable further analysis of the possible ways in which AI systems can be exploited for criminal activities. This research does not focus on developing a theory of malicious use and abuse of AI. With the typology presented in this paper, we hope to make the following contributions: a. Add to the emerging body of knowledge that maps types of malicious use and abuse of AI systems. To understand the main concepts, threat scenarios and possibilities is necessary to develop muchneeded preventive measures and proactive responses to such attacks. b. Help in establishing a shared language among and across different disciplines, especially between STEM disciplines and legal practitioners, as well as policymakers. Interdisciplinary research on the topic can reduce confusion caused by excessively technical or monodisciplinary language and aid in bridging existing gaps. c. Propose mitigation strategies, as well as demonstrating that a collective effort among government, academic and industry is

VII. RESULTS & DISCUSSIONS

The study aims to understand the threats posed by AI systems and develop mechanisms to protect society and critical infrastructures. It classifies AI systems into physical, psychological, political, and economic harm, and explores vulnerabilities in AI models and attacks like forgery. Collaboration between industries, governments, civil society, and individuals is crucial for developing knowledge and systems to address these challenges. Further research could use empirical methods and statistical analysis.

Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI

Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI

VIII. CONCLUSION

The proposed smart irrigation system, integrating IoT technology and machine learning, offers significant advancements over traditional irrigation methods. By utilizing real-time data from IoT sensors and advanced analytics, the system ensures precise irrigation management and optimizes water usage, leading to improved crop productivity. The system's ability to adapt to changing conditions and provide actionable insights enhances decision-making and resource management, promoting more sustainable and efficient agricultural practices. Overall, this innovative approach addresses the inefficiencies of conventional methods and supports modern, data-driven farming.

REFERENCES

[1] K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Arti_cial Intelligence. London, U.K.: Yale Univ. Press, 2021.

[2] D. Garcia, ``Lethal arti_cial intelligence and change: The future of international peace and security," Int. Stud. Rev., vol. 20, no. 2, pp. 334_341, Jun. 2018, doi: 10.1093/isr/viy029.

[3] T. Yigitcanlar, K. Desouza, L. Butler, and F. Roozkhosh, ``Contributions and risks of arti_cial intelligence (AI) in building smarter cities: Insights from a systematic review of the literature," Energies, vol. 13, no. 6, p. 1473, Mar. 2020, doi: 10.3390/en13061473. •

[4] I. van Engelshoven. (Oct. 18, 2019). Speech by Minister Van Engelshoven on Arti_cial Intelligence at UNESCO, on October the 18th in Paris. Government of The Netherlands. Accessed: Apr. 15, 2021. [Online]. Available: https://www.government.nl/documents/speeches/2019/ 10/18/speech-by-minister-van-engelshoven-on-arti_cial-intelligence-atunesco

[5] O. Osoba andW.Welser IV, The Risks of Arti_cial Intelligence to Security and the Future of Work. Santa Monica, CA, USA: RAND Corporation, 2017, doi: 10.7249/PE237.

[6] D. Patel, Y. Shah, N. Thakkar, K. Shah, and M. Shah, ``Implementation of arti_cial intelligence techniques for cancer detection," Augmented Hum. Res., vol. 5, no. 1, Dec. 2020, doi: 10.1007/s41133-019-0024-3

[7] A. Rodríguez-Ruiz, E. Krupinski, J.-J. Mordang, K. Schilling, S. H. Heywang-Köbrunner, I. Sechopoulos, and R. M. Mann, ``Detection of breast cancer with mammography: Effect of an arti_cial intelligence support system," Radiology, vol. 290, no. 2, pp. 305_314, Feb. 2019,

[8] J. Furman and R. Seamans, ``AI and the economy," Nat. Bur. Econ. Res., NBER, Cambridge, MA, USA,Work. Paper, 2018, doi: 10.3386/w24689.