International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

AI: A Double-Edged Sword in Cybersecurity- Threats and Defense Mechanisms

G Shirisha[1], Chandra Mouli Y[2], Vinodkumar G[3]

[1,2]Student, Computer Science and Engineering, Ballari Institute of Technology & Management
[3]Quality Assurance Engineer, Cybersecurity, OpenText

Abstract

Generally, when we hear something related to AI, it's always on how it helps us, how it reduces our workload or how it increases the efficiency of the task. But that's not always the case. This is just the one side of a coin. The other side of coin is not much popular in present times which in turn helps the people who misuse it. AI is a double-edged sword. As the sharpness of the sword increases, the better the sword will be. Since it's a double edged one, it hurts us too. In the exact same way as the AI evolves it helps people, but it also can be used to harm the people.

For example, if we see a soldier with a gun, we will feel safe but if we see the same gun with a stranger, we will be terrified. Here gun is akin to AI and normally the people only know about soldiers and pay no heed to the other situation. Hence in this paper we will be discussing about how AI can be used for wrong purposes.

This paper inspects the darker applications of AI, particularly the threats created by AI-driven cyberattacks. Our focus is to raise awareness of these obscure risks and stress the need for protections to prevent AI misuse. Specifically, we will cover the basics of cyberattacks, how AI makes them more effective, and a method to defend against them.

Keywords: Artificial Intelligence, Cybersecurity, AI-powered cyberattacks, Advanced Algorithms, Anomaly Detection, Automated Cyber Defense System

1. Introduction

A cyberattack is an intentional and harmful attempt by an individual or organization to breach another individual's or organization's information system. Usually, the attacker seeks some type of benefit from disturbing the victim's network. The first cyberattack is considered to happen on 1834 in France, where two thieves gained access to French telegraph system and stole financial market information [1]. It's funny to think that first cyberattack happened in an era without internet. But as the attacks started, so did the defenses. In 1940, Rene Carmille was the first person to be termed as the ethical hacker [1]. An ethical hacker is basically the one who chose the correct path to utilize his knowledge about systems or networks. We have seen that cyberattacks started way before internet was invented, it remarkably increased during late 20th century when the digital revolution started.

Cyber criminals were the first to adapt to these new emerging technologies and start stealing data or money [2]. One of the biggest cyberattack was Aadhaar data breach (2017-2018) in India. Between 2017 and 2018, India faced a series of breaches related to Aadhar, the nation's biometric identification system. The impact of these breaches was serious, as they potentially exposed the personal data of over 1.1 billion

individuals to identity theft and fraud [2]. At this point, cybercrime has affected the majority of internet users in some way. With the emergence of AI and its rapid development in recent years, cyber criminals will not let go this beautiful opportunity which will reduce their workload and help them in stealing. According to the Indian Cybercrime Coordination Centre, around 1,750 crore Rs was stolen by fraudsters between January and April in 2024 [3]. AI technology is increasing and will continue to increase. Instead of depending on others to resolve this issue, the least we can educate ourselves on this topic and try to avoid basic attacks. This following section focuses on basics about cyberattacks and how ai is used to enhance them.

AI powered Cyber-attacks is where attackers use the artificial intelligence for effectiveness of their attack. Attackers uses the AI tools for enhancing their attack like phishing attack, password cracking and identifying system vulnerabilities. AI is the art of creating machine that requires human intelligence.

For example, AI can send automated fake emails to get sensitive information by using natural language processing (NLP), Machine learning (ML). This involves where AI model is trained on Dataset of emails to understand structures [4]. The impact of this is AI will launch large scale operation than human could. AI powered cyberattacks are increasingly happening in industries like using phishing attacks, automated malware deployment, ransomware attack etc. This paper involves the how the attacks are happening with the help of AI and solution to prevent the dangerous attacks.

AI was introduced at Dartmouth college in the year of 1956.AI is art of creating machine that requires human intelligence. AI can mimic the human tasks such as problem solving, decision making etc. [5]. Where AI system can analyze the dataset to predict. Where AI is completely dependent upon the algorithms and model. Machine learning is core of AI it creates advanced algorithms and train the model by analyzing large dataset. AI is used in wide range of applications.AI consists of several subfields such as machine learning, deep learning, natural language processing. Usage of AI in cyberattacks is widely increasing. Where attacker can perform attack effectively with the help of AI.AI can used as both attacking and defensing system.AI can launch the attacks like phishing, malware, reconnaissance etc.AI can also prevent attack by using algorithms and training the models.

2. Literature Survey:

AI is evolving exponentially, which is both a good and bad thing as discussed above. The threats change constantly, and the methods used for countering threats change as well. Machine learning and deep learning are becoming popular in the cybersecurity domain and are labelled as game-changers. We have employed these methods in our methodology as well. In fact, every counter measure for ai driven cyberattacks require these two basic methods. These take large amounts of data in real time as inputs and find patterns and correlations. The algorithms then use these patterns to make predictions and decisions. The more data the algorithm consumes, the better it gets at finding patterns and trends. The existing research paper uses a mixed-methods approach, combining both qualitative and quantitative methods to evaluate AI applications in cybersecurity across various industries. It comprises a literature review of AI techniques (e.g., machine learning, NLP, deep learning) and examines case studies from diverse sectors like healthcare, finance, and technology. The methodology highlights historical data analysis, real-world applications, and challenges compares it to traditional methods, showing AI's efficiency in threat detection and response but also its reliance on quality data and human oversight [9].

Our approach focuses more on maximizing the use of AI for automated cyberattack simulations and defence systems, emphasizing real-time adaptability and predictive modelling for evolving threats. While

the cited methodology focuses broadly on applications and industry-specific implementations, ours provides a specialized, action-driven framework for integrating AI into active defence mechanisms, potentially filling gaps in proactive threat neutralization.

This means designing strategies that prioritize threat detection and neutralization rather than just responding to threats after they occur.

The existing approaches stresses on the development of adaptive, AI-based defence systems that proactively detect, prevent, and respond to cyber threats. By incorporating AI systems, it enables more dynamic and intelligent cybersecurity, as explained by Guembe et al. (2022) and Raimundo and Rosário (2021) [10]. Our approach aligns with this, focusing on preventing and neutralizing through AI. However, it's not entirely same, our methodology differs by integrating multiple components—advanced machine learning algorithms, behavioural anomaly detection, real-time autonomous dynamic responses, and a self-improving defence system—into one unified system. This approach is designed to handle a wide range of threats autonomously and adaptively, whereas other methodologies may focus more narrowly on specific attacks or require human intervention for training and adaptation.

3. Cyberattack Types:

Cyber-attacks are the attacks where attackers used to perform attack to steal personal information, exploit system vulnerabilities. Here are some of the attacks widely used by the attackers.

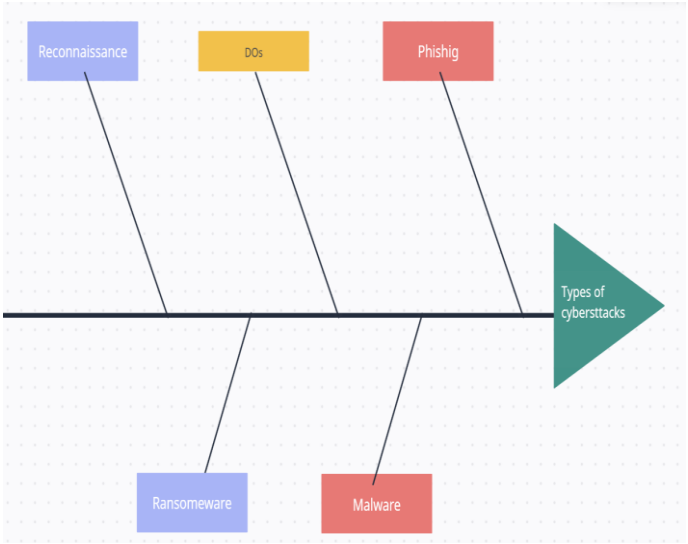

Figure 3.1: Types of Cyber-attack

- **Phishing:** Phishing is a type of cyber-attack in which attacker influence the people to reveals their sensitive information such as passwords, financial data, personal data.
  - AI powered phishing - AI is used to create personalized emails by analysing social media, online behaviour. These emails are harder to detect.

There are types of phishing such as vishing and smishing

a. **Voice Phishing (Vishing):** AI tools can create highly synthetic voices that mimic specific individuals.
b. **SMS Phishing (Smishing):** Attacker uses the AI to send deceptive text messages to trigger victims into sharing sensitive information.

- **Malware and Ransomware:** Malware is the software that cause the damage to system such as by deleting all the files or grant access to unauthorized access.

- AI Enhanced Malware - Malicious software that uses the AI to become smarter, adaptable and harder to detect. Traditional malware follows the fixed instruction but AI enhanced malware can learn, adopt, dynamically modify code and it can also choose the best files to attack. Polymorphic malware is malicious software constantly modifies itself.
- AI Driven Ransomware - AI can analyze networks and systems to identify high-value targets-such as important files, databases, or individual who are more likely to pay the ransom.

- **AI Deepfake based Attacks:** Attacker uses the AI to create fake audio, video and image to influence the people to commit the fraud this is very difficult to differentiate from the real ones. AI models can analyze recordings of person's voice and then replicate with high accuracy.
- **Reconnaissance:** Type of malware that locks or encrypts a victim's device or data and then demands payment to unlock it.
  - AI-powered reconnaissance: instead of manually looking for weakness AI can analyze publicly available information like social media and find out the weakness.
- **Botnet:** Network of infected computers that work together to carry out an attacker's goal.
- **Denial of Service (DoS):** Attacker make the website or service unavailable to user.
- **Crypto Jacking:** Attacker uses the someone else computer to mine cryptocurrency.
- **Advanced Persistent Threats (APT):** Where highly skilled attackers acquire unauthorized access to network which goes unnoticed for longtime.
- **Automated Attacks:** AI automated attacks are cyberattacking where artificial intelligence (AI) is used to carry out malicious actions automatically, without needing constant human input.
- **Credential Stuffing:** AI analyzes massive datasets of leaked credentials to automate login attempts, optimizing success rates.

4. Distribution of AI Application in Cybersecurity:

AI applications are widely used in cybersecurity to perform attack effectively and prevent attack [6]. Above figure represents the distribution of AI applications in cybersecurity.

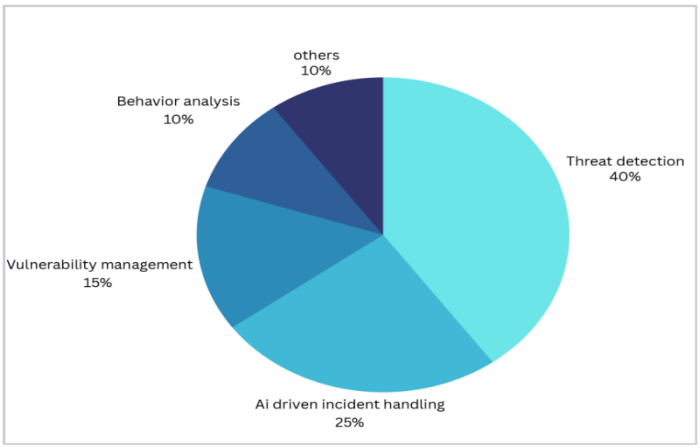

Figure 4.1: Distribution of AI applications in cybersecurity

a. **Threat Detection and Prevention:** AI is used to identify the threats by analyzing vast amount of data, user behavior, system behavior. And prevent attacks by using machine learning algorithms.[7]
b. **Behavioral Analysis:** AI will monitor the user activity to detect the suspicious attacks. For example, if user login behavior monitors if any deviation in the login time it will detect as attack.[8]

c. **AI Driven Incident Handling:** AI can respond to attack which made by the attacker.AI system can take the preprogrammed steps to respond to attack.
d. **Vulnerability Management:** AI can be able to be fixing weakness of computer system, software or network. It reduces the risk of attack by addressing these vulnerabilities.

5. Methodology

We propose this methodology, which consists of the following six steps.

Step 1: AI Automated Cyber Defense System

By analyzing few research papers, we have concluded that all of the mentioned methodologies include defense for the individual threats or attacks and also use methods such as user behavior anomaly detection or pattern detection and machine learning algorithms to prevent the attacks. This is capable of handling one type of threats.

Our methodology introduces the idea of integrating all of these i.e. Advanced machine learning algorithms, behavioral anomaly detection, automated dynamic response, self-improving defense mechanism to develop AI automated cyber defense system. All of these components are actively included and developed as integrated system. It's better to have a single model which is capable of doing all the things instead of having multiple software for multiple threats.

Automated cyber defense system is adaptable, that is if any new attack happens, it is capable of preventing by analyzing previous attacks or system vulnerabilities and takes immediate action in real time.

Step 2: Advanced Machine Learning Algorithms

This includes protecting AI models from the advanced attacks like data poisoning. This includes the continuous training of AI models to prevent the attacks, where both legal or valid data and manipulated data is fed into the AI model to detect the attacks.

Step 3: Behavior Anomaly Detection

This is an advance anomaly detection which goes beyond the normal or simple anomaly detection,
where it continuously learns from system behavior or user behavior. It collects the data from sources such as network logs, user activity etc. This machine learning model defines behavior of user, based on their data history. For example: user logs in between 6PM - 7 PM. If the model finds any deviation in user behavior, then it concludes it as an anomaly and asks some pre-defined questions or passwords which are known only to the user.

Step 4: Real Time Autonomous Dynamic Response

The current software requires human intervention to train the models. This should be integrated with real time, where it analyzes behavior anomaly detection and automatically prevents the attack.

Step 5: Self-Improving Mechanism

As the name suggests, it is self-improving or self-evolving model which learns to defend new attacks in real time. This can be done by reinforcement learning.

Step 6: Attack Prediction

We all know prevention is always better than cure. This applies in this case too, where predicting and preventing the attack from happening is far better than letting the attack happen. The model needs inputs such as threat intelligence feeds, history of the attacks, network traffic data, user activity logs, system vulnerabilities etc.

To build this AI automated cyber defense system (ACDS)requires AI engine that is capable of handling advance machine learning, behavior anomaly detection, real-time and dynamic response and self- improv-

ing mechanism. For this we need to create models by training with real time data.
AI automated cyber defense System.

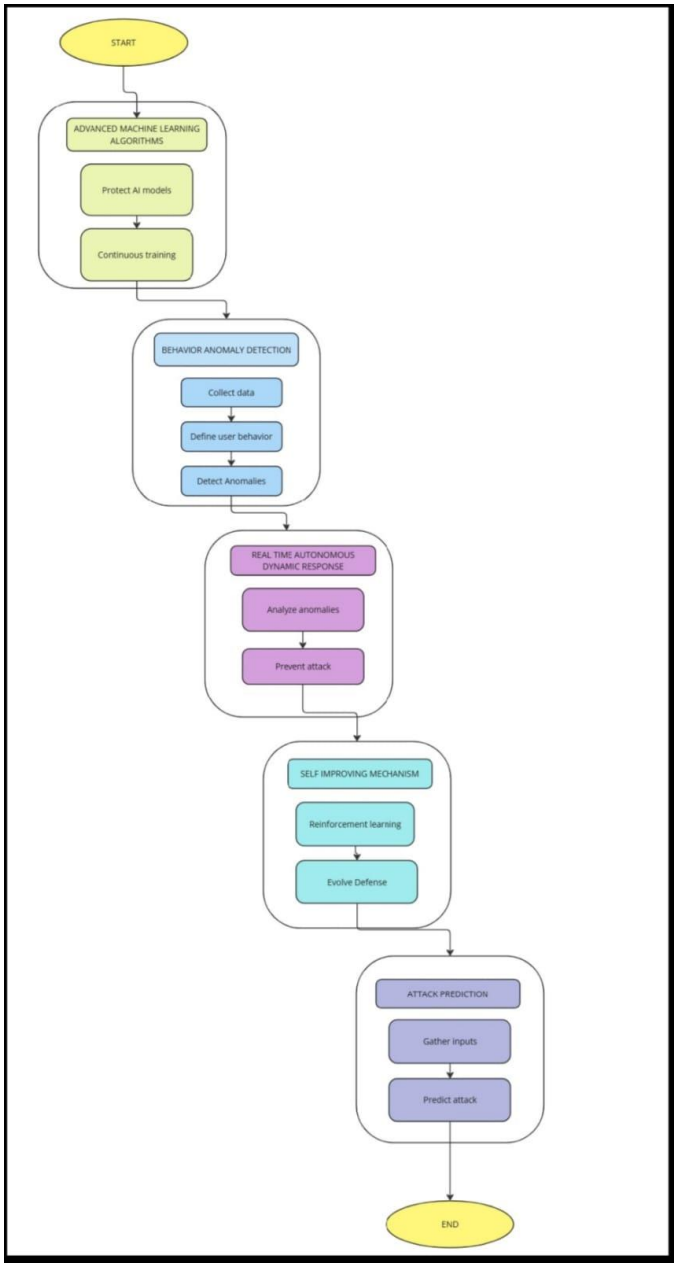

Figure 5.1: AI-Powered Cyber Defense Framework

6. Conclusion

Artificial Intelligence has dual nature.AI acts as major role in cybersecurity to perform attack effectively and for prevent the attack. By using AI attacker can perform the malicious attacks. Our methodology provides how we can prevent the existing and new attacks by using AI. AI automated cyber defense System is integrated system of advanced machine learning algorithms, anomaly detection, attack predicter, dynamic response and self-improving mechanism. By analyzing previous attacks, it can prevent the new attacks. AI automated cyber defense system is integrated, adaptive AI systems against emerging cyber threats.

References

1. https://www.monroeu.edu/news/cybersecurity-history-hacking-data-breaches
2. https://www.stldigital.tech/blog/10-biggest-cybersecurity-attacks-in-indian-history/
3. https://www.businesstoday.in/personal-finance/story/cybersecurity-for-the-everyday-user-simple-steps-to-protect-yourself-online-454097-2024-11-19#:~:text=According%20to%20the%20National%20Cyber,due%20to%20cyber%2Dcriminal%20activities.
4. Heiding, F., et al. (2024). Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns. *arXiv*. DOI: 10.48550/arXiv.2412.00586
5. Mijwil, M. M. (2015). *History of Artificial Intelligence*. Preprint. ResearchGate. DOI: 10.13140/RG.2.2.16418.15046.
6. Perception Point. (2024). AI in Cybersecurity: Applications and Use Cases. *Perception Point*. Retrieved from https://www.perception-point.io
7. Fortinet. (2024). AI in Cybersecurity: Enhancing Threat Detection and Prevention. *Fortinet*. Retrieved from https://www.fortinet.com
8. CrowdStrike. (2024). AI-Powered Behavioral Analysis in Cybersecurity. *CrowdStrike*. Retrieved from https://www.crowdstrike.com
9. Prince, N. U., Faheem, M. A., Khan, O. U., Hossain, K., Alkhayyat, A., Hamdache, A., & Elmouki, I. (2024). AI-Powered Data-Driven Cybersecurity Techniques: Boosting Threat Identification and Reaction. Nanotechnology Perceptions, 20(S10), 332–353. DOI: 10.13140/RG.2.2.22975.52644
10. Gudimetla, S. R., & Kotha, N. R. (2024). AI-Driven Cybersecurity: Enhancing Threat Detection and Response Strategies. International Research Journal of Modernization in Engineering, Technology, and Science, 6(5), May 2024. DOI: 10.56726/IRJMETS55883