International Journal of Research Publication and Reviews

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

Repercussion of ChatGPT in Cybersecurity

Dr. Preethi Ananthachari[a],Gurpreet Singh [b]

[a]*Assistant Professor,, Endicott College of International Studies, Woosong University, South Korea*
[b]*Student, Endicott College of International Studies, Woosong University, South Korea*

ABSTRACT

The ChatGPT large language model, developed by OpenAI, has significantly transformed the landscape of natural language processing and machine learning. This model has diverse applications ranging from chatbot development and question-answering to language generation. However, the extensive use of ChatGPT has led to concerns about its potential security implications. This research investigates the cybersecurity risks associated with ChatGPT and proposes methods to mitigate them. The study analyzes the security vulnerabilities in the model's training data, data processing, and language generation. Several solutions are suggested to address these issues, such as adversarial training, robustness testing, and implementing data privacy and security measures. The primary goal of this study is to raise awareness of the security implications of language models like ChatGPT and provide guidance on building secure and reliable AI systems.

Keywords: ChatGPT, Cybersecurity, OpenAI, Natural Language Processing (NLP), data privacy, Argumented Reality (AR), Virtual Reality (VR), Metaverse, Blockchain

Introduction

As we are entering into the modern world era of technology where humans are surrounded by technology from an automatic light lamp to coding by using Artificial Intelligence (AI). In this modern era where everything is online and shared continuously in every second in this whole world, we need to understand about the importance of data privacy. With the coming of ChatGPT professionals, working in every field felt the power of Artificial Intelligence (AI). Many job professionals, freelances and people from different fields have lost their jobs because of ChatGPT. According to the The Cybersecurity Hub , hackers have figured how to compromise ChatGPT to create phishing emails as well as Malware scripts. As ChatGPT is designed in such a way that when given any unethical command it will produce an output of error or saying that creating phishing emails of malware scripts is illegal, but hackers have found a new way to compromise it.

ChatGPT is created on a load of data and in such a way that it can learn from the humans by interacting and responding with the answers with the commands given by humans, ChatGPT is developed by an organization OpenAI which is an American artificial intelligence Institution for research. ChatGPT was released in the last of 2022 and gained more than 100 million users in January 2023.

Cybersecurity at one side is crucial for the prevention of Cyber Attack worldwide. Importance of Cybersecurity has increased over time with the introduction of technologies like Metaverse, Blockchain, Argumented and Virtual Reality. According to BBC news the importance of Cybersecurity is going to increase in the field of business as well.

Security risks

Cybersecurity researchers were able to create a malware script using ChatGPT even after the AI model was trained not to create. Compromising AI by finding a loophole in the AI system was just the part of researchers but at the same time this was also concerning because if the Cybersecurity researchers can compromise the system of AI, then what about the hackers which are present in worldwide. This is a concerning question that giving AI the power of securing the organization can be a complicated task. The similar incident also happened over time where Cybercriminals hacked into the chatbot and swamp it with malware commands, the cybercriminals had created their own bots that can gatecrashOpenAI’s GPT-3 API and amend the code. Once the code for amended, this malware bot was able to create malevolent content such as texts that can be used for phishing emails and malware scripts.

2. Distinct Approaches

ChatGPT can be used not only in creating phishing emails or malicious codes, but it is also capable of creating and helping in organizing Man-in-the Middle Attack (MITM), generate a list of most common password’s users are using in 2023 (Dictionary Attack), USB Rubber Duck Key scripts, Spoofing web pages, Denial of Service (DDoS) Attacks and much more. We should not forget that Artificial Intelligence (AI) abilities can be increased in a good way but can also be increased in a negative manner. One of the easiest ways by using Artificial Intelligence in the field of Cybersecurity can be Social Engineering Attacks. We should not forget that ChatGPT is built of load of data so asking for crucial information and surpassing this crucial information would be an easy task for a hacker, information like the number of employee working in a company, annual salary of particular person in that organization can be surpassed when ChatGPT will get more interactions with humans because there is no external server where the personal information's can be saved. There is only a main server where the people would be sharing about their day and about their future, so it would not be easy to get information about a single individual. Some of the Distinct Approaches are as follows:-

1. Compromising Email Account of the victim:- A hacker can get this crucial information in the very first case by compromising the email of the victim where all the questions asked by the user can be easily accessed by the hacker, now these questions can be an important part for the victim, personal information, or organization information in which the victim would be working.
2. Connecting your responsive page of ChatGPT with the victim:- There are various methods of performing such type of attacks in which the hackers can connect their web page as a hook for the victim's browser and one of the great and best examples of this is by using the Kali Linux tool BeEF.
3. Creating payload for the victim's system:- Using Metasploit payloads for compromising the systems are the easiest tasks used by hackers to get into the system and access the files.

As we can see that all the above methods are outside of ChatGPT and are the common methods a cybercriminal can use, but we should not forget that this can be dangerous if someone can find out that what we are interacting with an Artificial Intelligence (AI). Hackers can also compromise the server of ChatGPT for some time and can give permission to create malicious codes or much more as we have seen above.

3. Effective ways to prevent these attacks.

We need to understand the importance of our own data and what we are sharing while interacting with ChatGPT or Artificial Intelligence (AI). Some of the security measures we can follow are:-

1. Never share your personal information like name, date of birth, where you study, where you work, family information.
2. Never give commands about creating the page of Facebook, email, or any social website because ChatGPT uses cookies and caches and can save your personal information by extracting your profile from your browser.
3. Never ask to find someone using phone number or email or name or address of someone else because this night end up sharing someone else details to Artificial Intelligence and this may not end up good.
4. Never share your future plan or your daily routine with ChatGPT because this is your personal information and can be taken in a wrong manner in future by AI.

4. Conclusion

The data is most important asset in today's modern world and when we talk about Artificial Intelligence (AI) we can understand that it has the ability to improve itself by more interactions with more humans and also cybercriminals can use ChatGPT for easing the work in compromising systems and therefore we need to understand about the working with Artificial Intelligence for advantages by learning something new not by disclosing our own personal information which can be used by Artificial Intelligence to check us through from every social website from everywhere on the internet and can provide our personal information to the cybercriminals of normal people. We have discussed various points through which OpenAI'sChatGPT can share our information and how this information can become so important for hackers and cybercriminals, and we have also discussed how one can be secure and make themselves protect against these.

References

[1] Winder, D. (2023, February 3). *Does ChatGPT Pose A Cybersecurity Threat? Here's The AI Bot's Answer*. Forbes. https://www.forbes.com/sites/daveywinder/2023/02/03/does-chatgpt-pose-a-cybersecurity-threat-heres-the-ai-bots-answer/?sh=7c040833505d

[2] Zorz, M. (2023, January 24). *ChatGPT is a bigger threat to cybersecurity than most realize - Help Net Security*. Help Net Security. https://www.helpnetsecurity.com/2023/01/26/chatgpt-cybersecurity-threat/

[3] Cyber Security Hub. (2023, January 11). *Cybercriminals are using ChatGPT to create malware*. https://www.cshub.com/malware/news/cybercriminals-are-using-chatgpt-to-create-malware

[4] Limited, I. (2023, February 3). *ChatGPT presents new risks – here are five things you can do to mitigate them*. https://www.infosys.com/insights/cyber-security/new-risks.html

[5] Cyber Security Hub. (2023b, February 14). *Can ChatGPT be used for cyber attacks?* https://www.cshub.com/attacks/articles/chatgpt-cyber-attack-threat