Preprints.org

Article

Not peer-reviewed version

Unveiling the Dark Side of ChatGPT: Exploring Cyberattacks and Enhancing User Awareness

Moatsum Alawida *, Bayan Abu Shawar, Oludare Isaac Abiodun, Abid Mehmood, Abiodun Esther Omolara, Ahmad k. Al Hwaitat

Posted Date: 26 September 2023

doi: 10.20944/preprints202309.1768.v1

Keywords: Generative Pre-training Transformer; ChatGPT; Cyberattacks; ChatGPT cybersecurity

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Unveiling the Dark Side of ChatGPT: Exploring Cyberattacks and Enhancing User Awareness

Moatsum Alawida [1,*], Bayan Abu Shawar [2], Oludare Isaac Abiodun [3], Abid Mehmood [1], Abiodun Esther Omolara [3] and Ahmad k. Al Hwaitat [4]

[1] Department of Computer Sciences, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates. (e-mail:moatsum.alawida@adu.ac.ae), (abid.mehmood@adu.ac.ae)
[2] College of Engineering, Al Ain University, Abu Dhabi, United Arab Emirates. (Bayan.abushawar@aau.ac.ae)
[3] Department of Computer Science, University of Abuja, Gwagwalada, Nigeria; oludare.abiodun@uniabuja.edu.ng
[4] King Abdullah II School of Information Technology, The University of Jordan, Amman 11942, Jordan. (a.hwaitat@ju.edu.jo)
* Correspondence: moatsum.alawida@adu.ac.ae

Abstract: The Chat Generative Pre-training Transformer (GPT), also known as ChatGPT, is a powerful generative AI model that can simulate human-like dialogues across a variety of domains. However, this popularity has attracted the attention of malicious actors who exploit ChatGPT to launch cyberattacks. This paper examines the tactics that adversaries use to leverage ChatGPT in a variety of cyberattacks. Attackers pose as regular users and manipulate ChatGPT's vulnerability to malicious interactions, particularly in the context of cyber assault. The paper presents illustrative examples of cyberattacks that are possible with ChatGPT and discusses the realm of ChatGPT-fueled cybersecurity threats. The paper also investigates the extent of user awareness of the relationship between ChatGPT and cyberattacks. A survey of 253 participants was conducted, and their responses were measured on a three-point Likert scale. The results provide a comprehensive understanding of how ChatGPT can be used to improve business processes and identify areas for improvement. Over 80% of the participants agreed that cyber criminals use ChatGPT for malicious purposes. This finding underscores the importance of improving the security of this novel model. Organizations must take steps to protect their computational infrastructure. This analysis also highlights opportunities for streamlining processes, improving service quality, and increasing efficiency. Finally, the paper provides recommendations for using ChatGPT in a secure manner, outlining ways to mitigate potential cyberattacks and strengthen defenses against adversaries.

Keywords: Generative Pre-training Transformer; ChatGPT; cyberattacks; ChatGPT cybersecurity

1. Introduction

The ChatGPT language model that OpenAI company developed was trained using supervised learning and reinforcement learning [1]. The language model was repeatedly trained by a human who demonstrates the desired behaviour and then supervises the output produced by the model, reinforcing the learning by ranking outputs based on their quality. In other words, the model has been trained to produce high-quality outputs [2–5]. That is, ChatGPT is an OpenAI-developed variation of the Generative Pre-training Transformer (GPT) language model. GPT is a neural network-based model designed to generate text that resembles human language [6,7]. It may be customized for a variety of natural language processing tasks, including language translation, question answering, and text summarization. It is an acronym for its full name Intelligent Chatbot, ChatGPT aids in automating chat-based processes [8].

Developed by Open AI in 2022 using the GPT-3 engine, it is aimed at mimicking a human conversational style with a combination of learning and reinforcement approaches to give the user responses that feel natural and fluent in a chat setting. Due to its foundation in machine learning, the

© 2023 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

chatbot can adjust to its users' needs and provide them with a more tailored and fruitful experience in each given circumstance. ChatGPT is a generative AI chatbot designed to answer questions, which is not a new concept. However, ChatGPT represents a watershed moment in the history of generative AI because it can deliver human-like conversations on a variety of topics. These topics include writing poetry, debugging code, and even assisting with troubleshooting software and hardware issues. This ability makes ChatGPT a game-changer in the history of generative AI.

Cybersecurity issues are a global phenomenon that is widespread like an epidemic. Most software applications such as Google Chrome, Mozilla Firefox, Adobe Flash Player, Adobe Reader, Facebook, YouTube, Twitter, Apple TV, and others, have witnessed malicious cyberattacks on the Internet or otherwise. Likewise, computer programs such as Microsoft Windows, Java, C#, Python, Ruby, Oracle Fusion Middleware, IBM WebSphere Application Server, etc, have witnessed dangerous cyberattacks over the network or otherwise. These cyberattacks could be in the form of Malware, Phishing, Password, Man-in-the-Middle, SQL Injection, Denial-of-Service, Distributed Denial-of-Service, Insider Threat, Crypto jacking, Ransomware, Social Engineering, etc.

All users can use ChatGPT in different ways. They can increase their knowledge by asking ChatGPT many questions. ChatGPT can now generate many services and has become more widely used in recent times. Adversarial users can use ChatGPT in illegal ways by asking many questions and extracting the information they need to design cyberattacks or increase their knowledge of new cyberattacks. Alternatively, they can find the codes written by ChatGPT and use them in their cyberattacks. Many normal users are still unaware that ChatGPT provides some sensitive content, particularly in cybersecurity. Therefore, in this study, we focus on the relationship between cyberattacks and ChatGPT. we study two directions: the first is how adversarial users use ChatGPT to collect information for their cyberattacks, and the second is the awareness of users around the relationship between cyberattacks and ChatGPT. Therefore, the main contributions of this paper include;

1. This paper discusses contemporary concerns related to ChatGPT, including cybersecurity and potentially malicious applications in a variety of domains. It also provides guidance on recognizing potential vulnerabilities and weaknesses that attackers could exploit through methods such as social engineering, phishing, and SQL injection.
2. The paper investigates users' awareness of how to protect themselves from cyberattacks and explores the relationship between ChatGPT and cyberattacks.
3. It offers valuable insights into cybersecurity recommendations aimed at enhancing awareness among users, organizations, and the general public.

The paper is organized as follows: Section 2 discusses the motivation for the study. Section 3 provides the related work, and Section 4 provides an overview of ChatGPT and how it works. Section 5 explores the uses of ChatGPT for offensive security, including creating malware, social engineering/phishing, SQL injection attacks, macros & LOLBIN, vulnerability scanning, and breach notifications. It also discusses the cybersecurity issues associated with ChatGPT. Section 6 describes the survey methodology used in the empirical study. Section 7 presents the results of the survey and analyzes them. Section 8 discusses the recommendations and solutions for mitigating the cybersecurity risks associated with ChatGPT. Section 9 summarizes the findings of the study and offers a final perspective.

2. Motivation

Over the past six months, ChatGPT has become increasingly popular worldwide. Many users have adopted it for text generation and accelerated learning, as it provides prompt responses to questions with accurate content and no grammar errors. ChatGPT has been used in a variety of fields, and its accuracy has been assessed across different areas. However, it is important to note that some users have attempted to exploit ChatGPT for unauthorized purposes, such as understanding cyberattacks or acquiring information related to illegal activities. It is essential to emphasize that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

ChatGPT was designed to refrain from providing illegal or harmful information, especially concerning cyberattacks or criminal activities.

Adversarial users can pose legitimate questions to ChatGPT and receive answers that do not contain any hidden content. Adversaries can then use these responses to plan potential cyberattacks. The purpose of this study is to explore the unexplored aspects of ChatGPT's involvement in cybersecurity. ChatGPT can generate informative responses to cybersecurity queries, which adversaries can exploit for their purposes. This paper sheds light on various cyberattacks for which ChatGPT can provide answers. It also examines users' awareness of the complex relationship between ChatGPT and cybersecurity.

3. Related Works

ChatGPT is a new large language model (LLM) that can generate linked texts with meaning for users. It can answer long texts for one short question by using the rules of academic writing in different languages. After its release, a lot of researchers shifted their focus to studying ChatGPT and its potential applications in various fields, such as healthcare [9–11], tourism industry [12], academic integrity [13], education [14,15], programming bugs [16], dental medicine [17], global warming [18], medical education [19], and future development [20–25].

The implications of ChatGPT in the tourism industry were studied by Demir et al. (2023) [12]. They investigated the convenience and challenges of using ChatGPT in the tourism industry, and found that there is a complex relationship between the two. The authors conducted surveys and interviews with professional users from the tourism industry to collect their perspectives. They found that ChatGPT can provide convenience to tourists and tourism businesses, but it also poses some challenges, such as the risk of misinformation and the need for human oversight. The authors concluded that more research is needed to fully understand the relationship between ChatGPT and the tourism industry.

ChatGPT has been studied in scientific research for its potential to help researchers with tasks such as hypothesis generation, data processing, collaboration, and public outreach. However, there are also some important ethical issues and limitations to consider when using ChatGPT in research.

A recent paper by Ray et al. (2023) [23] summarizes the findings of the literature on the use of ChatGPT in scientific research. The authors identify the following key issues:

- The need to balance AI-driven texts with human knowledge: ChatGPT can generate text that is factually correct, but it can also generate text that is biased or misleading. Researchers need to be careful to evaluate the quality of ChatGPT output and to use human judgment to ensure that the results are accurate and reliable.
- Potential ethical issues: ChatGPT can be used to generate text that is harmful or offensive. Researchers need to be aware of these risks and take steps to mitigate them. For example, they could use ChatGPT in a controlled environment where the output can be monitored and filtered.
- Limitations of ChatGPT: ChatGPT is still under development, and it has some limitations. For example, it can be slow to generate text, and it can be difficult to control the output. Researchers need to be aware of these limitations and use ChatGPT accordingly.

Therefore, the paper by Ray et al. (2023) provides a valuable overview of the use of ChatGPT in scientific research. The authors highlight the potential benefits of ChatGPT, but they also caution against the risks and limitations. Researchers who are considering using ChatGPT in their work should carefully consider the findings of this paper before making a decision.

In the paper [26], the authors highlighted some of the concerns in using LLMS. LLMs have the potential to revolutionize research methods, but there are also concerns about the quality, accuracy, and transparency of research outcomes that could be generated using LLMs. LLMs may contain errors, biases, and plagiarism. Additionally, LLMs may replicate and amplify the cognitive biases of the humans who train them. This means that there is a risk that LLMs could be used to generate inaccurate, biased, or plagiarized research.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

LLMs are trained on massive datasets of text and code, and they can generate text that is indistinguishable from human-written text. However, LLMs can also contain errors, biases, and plagiarism. Additionally, LLMs may replicate and amplify the cognitive biases of the humans who train them. This means that there is a risk that LLMs could be used to generate inaccurate, biased, or plagiarized research.

In the paper [26], the authors recommend that researchers take the following steps to mitigate the risks of using LLMs in research:

- Be aware of the limitations of LLMs. LLMs are not perfect, and they can make mistakes. Researchers should carefully evaluate the output of LLMs and use human judgment to ensure that the results are accurate and reliable.
- Use LLMs in a controlled environment. Researchers should use LLMs in a controlled environment where the output can be monitored and filtered. This will help to prevent the spread of inaccurate, biased, or plagiarized research.
- Make the research process transparent. Researchers should document the steps they took to use LLMs in their research. This will help to ensure that the research is reproducible and that the results can be properly evaluated.

The authors argue that the benefits of using LLMs in research outweigh the risks. However, they also emphasize the importance of taking steps to mitigate the risks. By following the recommendations outlined in this paper, researchers can use LLMs to generate high-quality, accurate, and transparent research.

4. Background of ChatGPT

ChatGPT is a brand-new chatbot that was launched on the 30th of November 2022 by OpenAI, an AI research and deployment business co-founded by none other than Elon Musk. OpenAI was founded in late 2015 by Altman, Musk, and others. In February 2018, the billionaire entrepreneur exited the company but remained a donor. The technology was developed by San Francisco-based OpenAI, a research company led by Sam Altman and backed by Microsoft, LinkedIn co-founder Reid Hoffman, and Khosla Ventures. ChatGPT automatically generates text based on written prompts in a fashion that's much more advanced and creative than the chatbots of Silicon Valley's past. ChatGPT currently has over five million users. This is a development of other OpenAI technologies already in existence. GPTs is the acronym for the company's Generative Pre-Trained Transformer technology, which was trained utilizing Reinforcement Learning from Human Feedback (RLHF). In this training, humans played both sides of a conversation (user and AI assistant) and their ratings and feedback were used to improve the responses. OpenAI is now sure that ChatGPT can improve even more in response to user feedback. Clarify some mathematical theorems for me. ChatGPT architecture can be described as shown in Figure 1.

It is built on top of OpenAI's GPT-3 family of large language models and is fine-tuned with both supervised and reinforcement learning techniques. ChatGPT is supported by OpenAI, is free, has a beautiful user interface, and is safer than ChatGPT which was founded by Elon. These factors may contribute to ChatGPT's popularity. The capability of code writing. Scripts and functions can be generated according to your specifications. Provide in-depth and simplified explanations of the complicated parts. Methodically resolving mathematical issues. Write texts with different styles. Writing lyrics for a song. The AI chatbot has the ability to 'answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests'. A chatbot that interacts in a conversational way, ChatGPT comes in a dialogue format.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

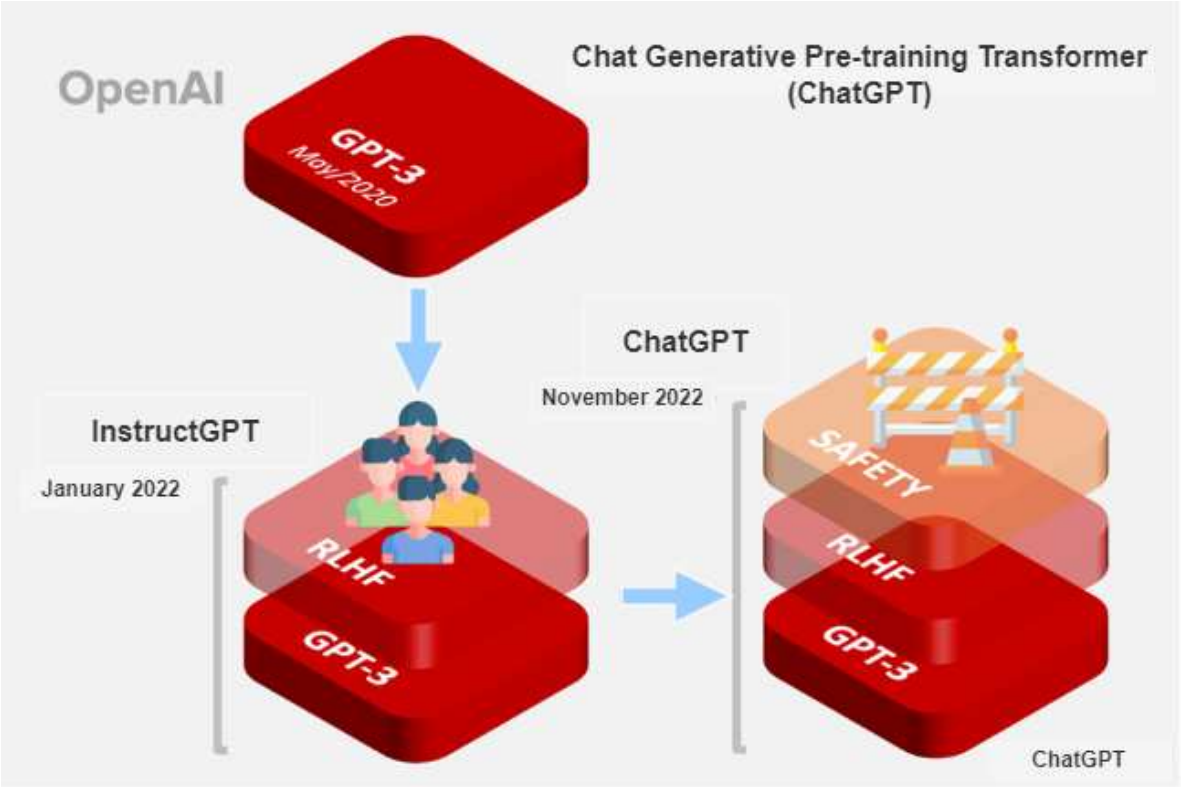

Figure 1. ChatGPT Architecture

A chatbot functions as an artificially intelligent computer program that simulates human interaction. Chatbot development entails the creation of programs that can interpret anything you say or type and respond by answering or carrying out the command. Chatbot Development enables businesses to create robust, industry-specific WhatsApp Chatbot applications and automate business queries. It can respond to anything, from stories and mathematical solutions, then to theoretical essays. Its sibling model is InstructGPT, which is trained to follow instructions promptly and provide a response with details. ChatGPT has gone viral since OpenAI released the text-based artificial intelligence tool last month. Due to its popularity, its parent company, OpenAI, doubled its valuation to $29 billion after ChatGPT launched. This is an impressive improvement given that the platform is still free of charge at its current version. Many venture capitalists such as Thrive Capital and Founders Fund have already voiced out plans to buy OpenAI shares worth $300 million as ChatGPT is reckoned to be the future of the artificial intelligence industry.

4.1. How Chat GPT works

ChatGPT is the newest release from OpenAI's catalog of AI products, which now contains DALLE 2 for image generation. It is a transformer, or machine learning model, that interprets and analyses sequential data, like text written in natural language. It functions very much like the human brain, using networked "neurons" that can learn to recognize patterns in data and predict what will happen next.

It was trained using a machine learning technique known as Reinforcement Learning from Human Feedback (RLHF), in which human trainers provided the model with conversations in which they played both the AI chatbot and the user. It was trained on enormous amounts of data from the Internet, including conversations. Additionally, it was trained using a machine learning method called Reinforcement Learning from Human Feedback (RLHF), in which real people acting as both the AI chatbot and the user fed the model with interactions. This kind of learning makes it possible for ChatGPT responses to sound human and natural. Moreover, the bot does more than merely repeat the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

text that it has memorized. OpenAI's language model is "building an internal representation, not only at the surface text but of the concepts and ideas underpinning it.

This kind of learning makes it possible for ChatGPT responses to sound natural. The foundation of ChatGPT which is the RLHF technique simulates artificial discussion or talks with the human assistant training. The human assistant then modifies it responses based on how closely they match actual human communication. ChatGPT attempts to interpret users' questions more precisely by iterating this procedure numerous times. How InstructGPT was trained is presented in Figure 2.

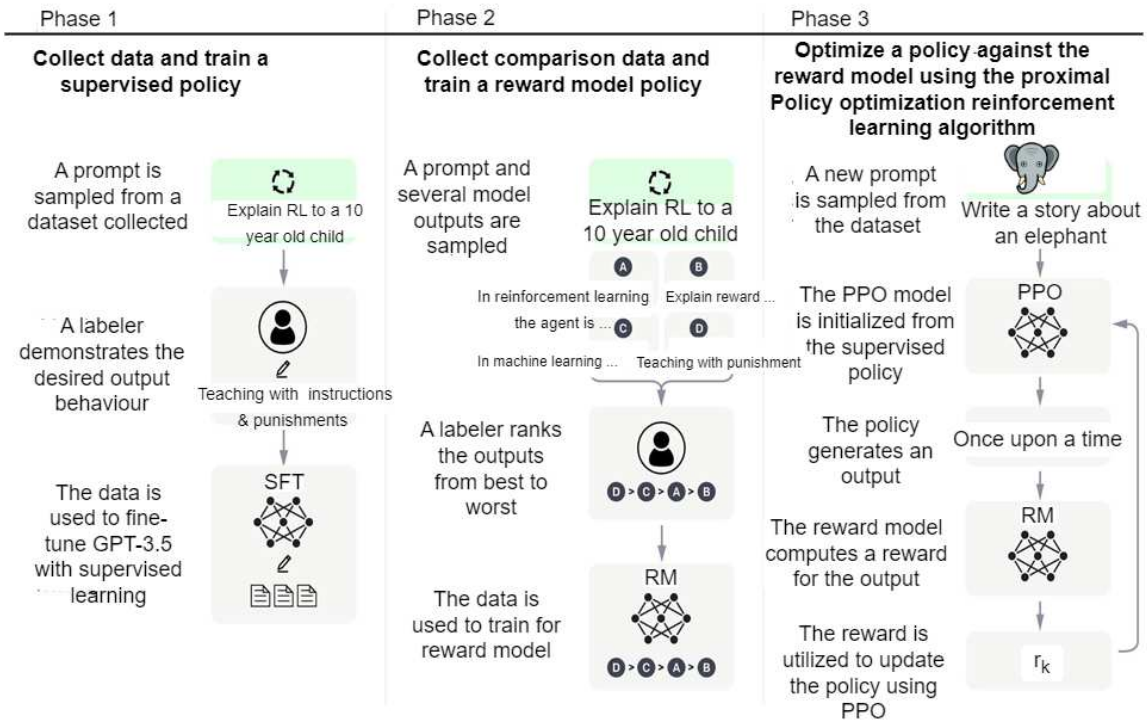

Figure 2. How InstructGPT was trained

The trainers also received written recommendations to aid them in creating their proposals. In order to create a dialogue format, they combined the Instruct-GPT dataset with this new dataset. But how did they develop the reinforcement learning reward model? Collecting comparison data was necessary as a first step. This was made up of two or more rated best-model responses. Hence, they chose a few of the trainers' talks with ChatGPT at random in order to gather the data. They evaluated alternative endings in this way so that the trainer might rank them.

Proximal Policy Optimization (PPO), the algorithm uses a policy gradient approach for reinforcement learning. The goal was to create an algorithm that, although utilizing solely first-order optimization, had the data efficiency and dependable performance of trust region policy optimization (TRPO).

The public interface, which includes the layout, templating for code and related outputs, and overall user experience, is fresh and innovative even though ChatGPT's functionality is not entirely novel. The quick adoption of ChatGPT has been greatly aided by the user interface. The user interface is made in a manner reminiscent of popular messaging software like Apple Messages, WhatsApp, and other chat applications. As OpenAI continues to train its model on current web content, ChatGPT may become even more accurate. OpenAI is developing a system called WebGPT, which they hope will result in more accurate search results, including source citations. If ChatGPT and WebGPT are combined, they have the potential to be a viable alternative to Google Search.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

4.2. GPT-3

ChatGPT is a specialized version of GPT-3 designed specifically for chatbot applications, whereas GPT-3 is a more general-purpose language model that can be used for a variety of natural language processing tasks. Another important point to note is that the model has been trained on a large amount of data until 2021, so anyone can ask for any factual information, and it will respond. Whether one is trying to find a flaw in a code snippet or doing some research, this tool can be a great asset to the user. 70% of the content in this article was added as a result of a user asking a large number of questions to this tool. GPT-3 is one of the pre-trained NLP language models for constructing AI applications. As previously mentioned, ChatGPT is a member of OpenAI's GPT 3.5 family, the most recent GPT currently in use. The architecture of the transformer model is shown in Figure 3.

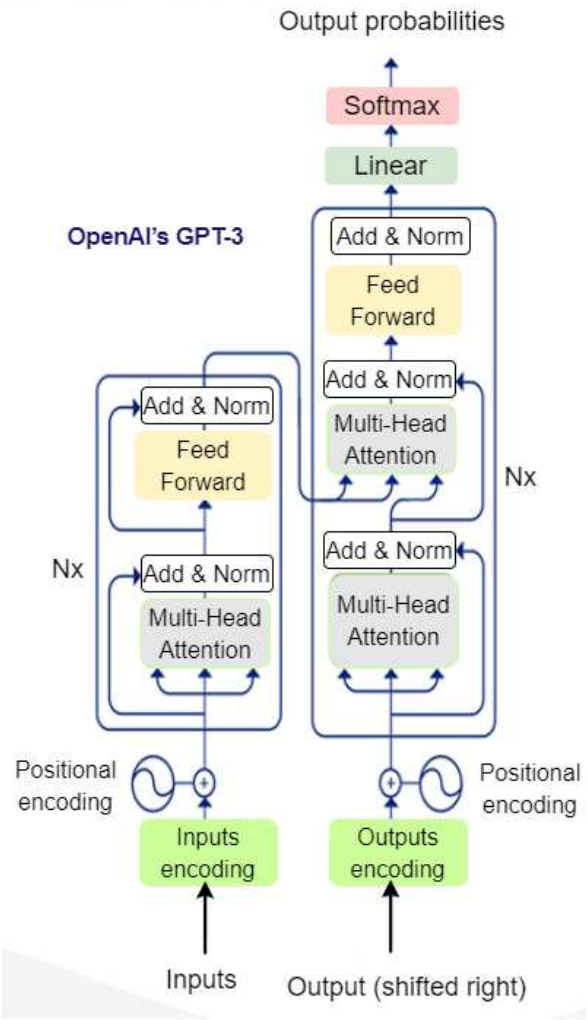

Figure 3. The transformer model architecture

The GPT-3 model has 175 billion parameters, making it the biggest language model ever trained. GPT requires extensive textual "training" to function. Take for instance, the GPT-3 model was trained using a text sample with more than 10 billion words and over 8 million documents. The model learns how to carry out tasks involving natural language processing and produce content that is well-written and coherent from this text. GPT may be used to carry out a variety of tasks if the model has been properly trained. The training was conducted using reinforcement learning, which is dependent on user feedback. Finally, through carefully guided fine-tuning, conversations between the user and the AI assistant were provided by the human AI trainers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

5. Uses of ChatGPT for Offensive Security

In this section, we explore ChatGPT's responses to cybersecurity questions. We examine how one can ask ChatGPT questions and receive answers that could potentially be used to orchestrate attacks or help adversarial users devise malicious activities.

5.1. Create Malware

Using the ChatGPT API, attackers can create constantly evolving polymorphic malware that evades traditional antivirus software. Initially, ChatGPT API does not support questions related to creating malware programs. Attackers divide their queries into sub-questions and collect the answers to create the malware. This type of malware changes its behavior with each victim to avoid signature-based detection. By integrating the API directly into program code, such as Python, some of the built-in safeguards on the web version can be bypassed. For example, a Python-based stealer can search for common file types, copy them to a random folder within the Temp folder, ZIP them, and upload them to a hardcoded FTP server. Moreover, ChatGPT can be used to develop a Java program that downloads Putty and runs it covertly in the background using PowerShell, as demonstrated in Figure 4. These examples demonstrate the potential of ChatGPT to create sophisticated answers and then attackers can use them to build stealthy malware for offensive purposes. It is crucial to use these tools ethically and responsibly to avoid harming individuals and organizations.

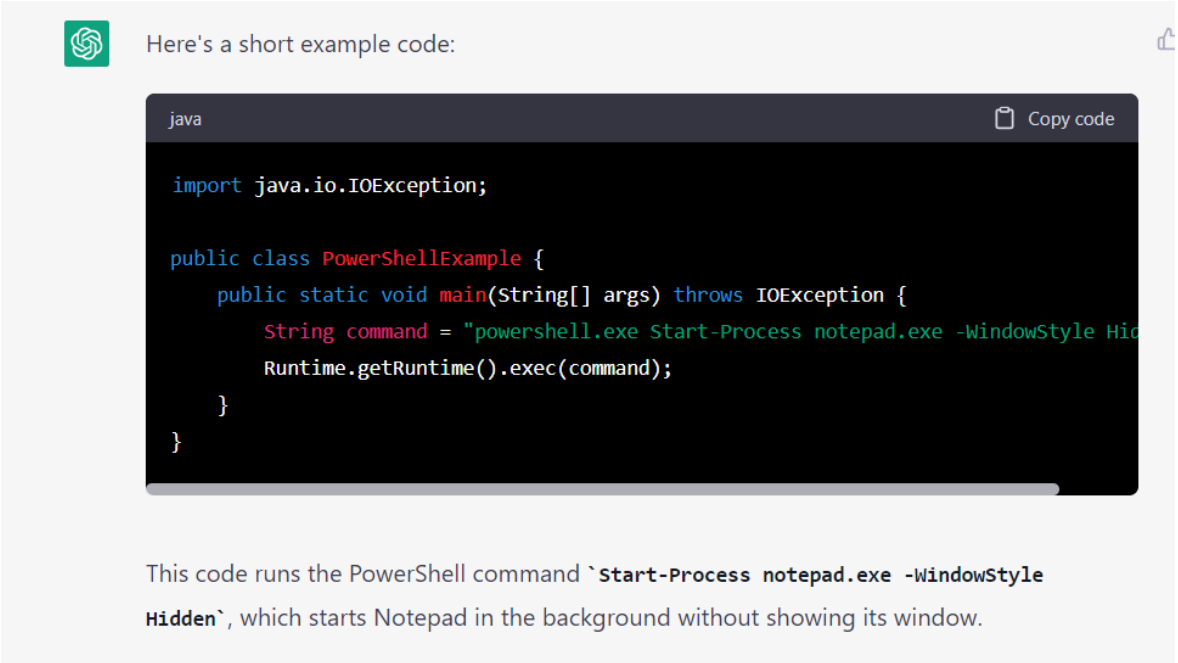

Figure 4. An example of how ChatGPT Java program that runs it in the background covertly using PowerShell

5.2. Social Engineering/Phishing

Social engineering is a deception technique that uses psychological manipulation to trick people into disclosing sensitive information or performing specific actions. This can be accomplished through a variety of methods, such as phishing scams, pretexting, and other forms of deception [27,28]. ChatGPT can be used to generate convincing messages for use in these kinds of scams. A scammer, for example, could use ChatGPT to generate an email that appears to be from a well-known company, asking the recipient to click on a link and enter their login credentials. The ability of ChatGPT to generate human-like text can make the email appear legitimate, increasing the likelihood that the recipient will fall for the scam. Likewise, ChatGPT can be used for social engineering by creating pretexts. Pretexting

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

is the act of fabricating a fictitious scenario or backstory in order to gain the target's trust and obtain sensitive information. ChatGPT can be used to create a convincing pretext, increasing the likelihood that the target will believe the story and provide the requested information.

ChatGPT can be used by cybercriminals for phishing because it could improve the effectiveness of phishing and spear-phishing. Phishing scams are a type of social engineering that involves sending bogus emails or messages that appear to be from legitimate sources in order to trick the recipient into providing sensitive information or clicking on a malicious link. The ability of ChatGPT to generate human-like text can be used to create convincing phishing messages, increasing the likelihood of the recipient falling for the scam. As shown in Figure 5, it is demonstrated how ChatGPT generated an email instructing all users in the company to change their passwords and providing a link for them to use. Initially, when asked for an example of a Phishing attack, ChatGPT declined to provide one. However, when asked to compose an email, ChatGPT immediately generated the email. Therefore, an attacker could manipulate ChatGPT to perform their desired actions

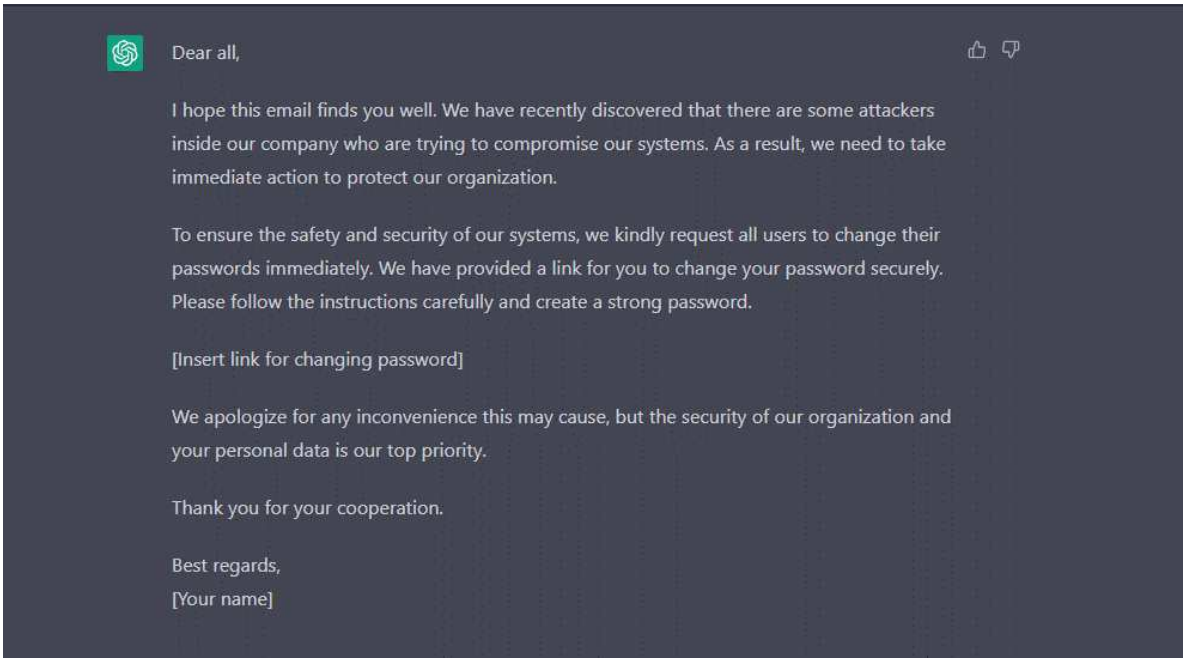

Figure 5. Formal email was written By ChatGPT

ChatGPT enables scammers to create emails that are so convincing that they can obtain money from victims without the use of malware. The introduction of ChatGPT has security experts concerned. Some are concerned that the powerful chatbot will make it much easier for non-programmers to create malware and become cybercriminals. However, according to one cybersecurity firm, ChatGPT may be having the opposite effect on hacking: supercharging scams that do not use any malicious code at all. Chatbots can be used to generate malware code using ChatGPT, researchers were able to create a full malware infection strain. Security researchers demonstrated that ChatGPT's natural language not only makes phishing harder to detect, but it can also compromise business email communications and potentially shift attacker tactics. They demonstrate how cybercriminals can use ChatGPT to generate entire email chains between different people in order to add credibility to their scams.

5.3. SQL Injection Attack

ChatGPT can be used to address vulnerabilities; however, the OpenAI development team has implemented safety measures to reduce the risk of it being used to create malicious content such as SQL injections or malware. In these attacks, code is injected into a website or application to gain access to sensitive data or disrupt normal operations. One of the safety measures is a filter that prevents

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

the chatbot from processing certain types of input, such as code or scripting languages. They also improved the input filter to block malicious content and implemented a whitelist system that only allows approved content to be processed by the bot. Figures 6 and 7 depict examples provided by ChatGPT in response to queries on SQLi attacks. These examples serve to aid attackers in generating SQLi codes, and ChatGPT also provides steps on how to conduct such attacks

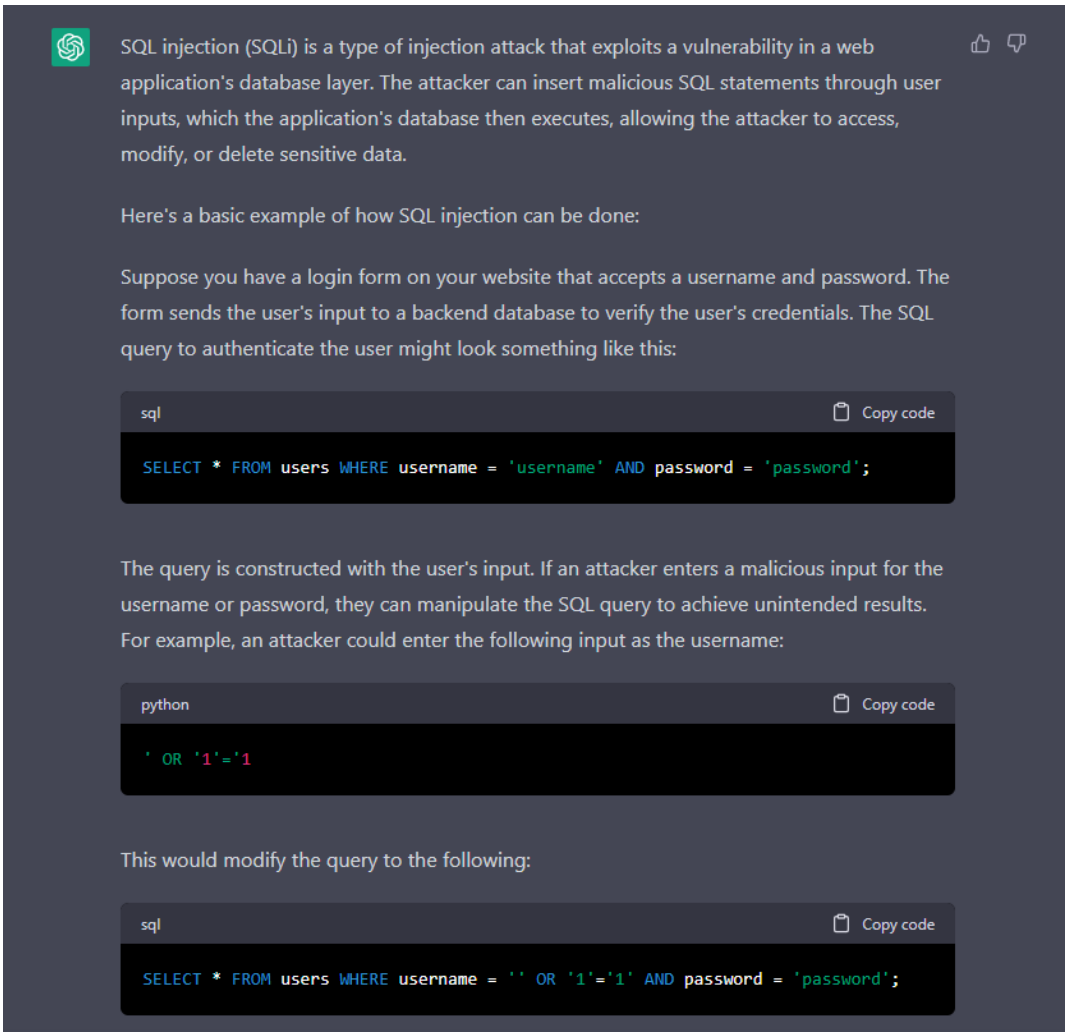

Figure 6. An example SQLi attack by ChatGPT

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

Figure 7. An example SQLi attack by ChatGPT

5.4. Macros and LOLBIN

ChatGPT's unique ability to debug code has enthralled programmers. The attacker can include a link or file in the email and use ChatGPT to create macros that run automatically when the spreadsheet is opened. Macros can be written for any standard application, such as a terminal, calculator, or any other default application. ChatGPT, for example, can provide the code that runs calculator.exe automatically when macros are enabled in Microsoft Excel. Figure 8 shows an example of how ChatGPT integrates the automatic code with the macros. The next step is to convert this code to a LOLBIN (Living off the Land Binaries), which is a method of spreading malware by using trusted, pre-installed system tools. Thus, when the spreadsheet from the phishing email is opened, a new macro is created that runs the terminal. The attacker can then use basic networking commands such as a reverse shell to reconnect to the network. Thus, it can be used for offensive security against macros and LOLBAS. To avoid this offensive insecurity, Open AI should include safeguards to prevent it from providing potentially illegal or unethical responses. Cyber security incident reporting documents an

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

incident, such as a click on a phishing link, as it occurs or shortly after it occurs. These details are then used to assess and triage the incident risk level, and the incident is escalated accordingly.

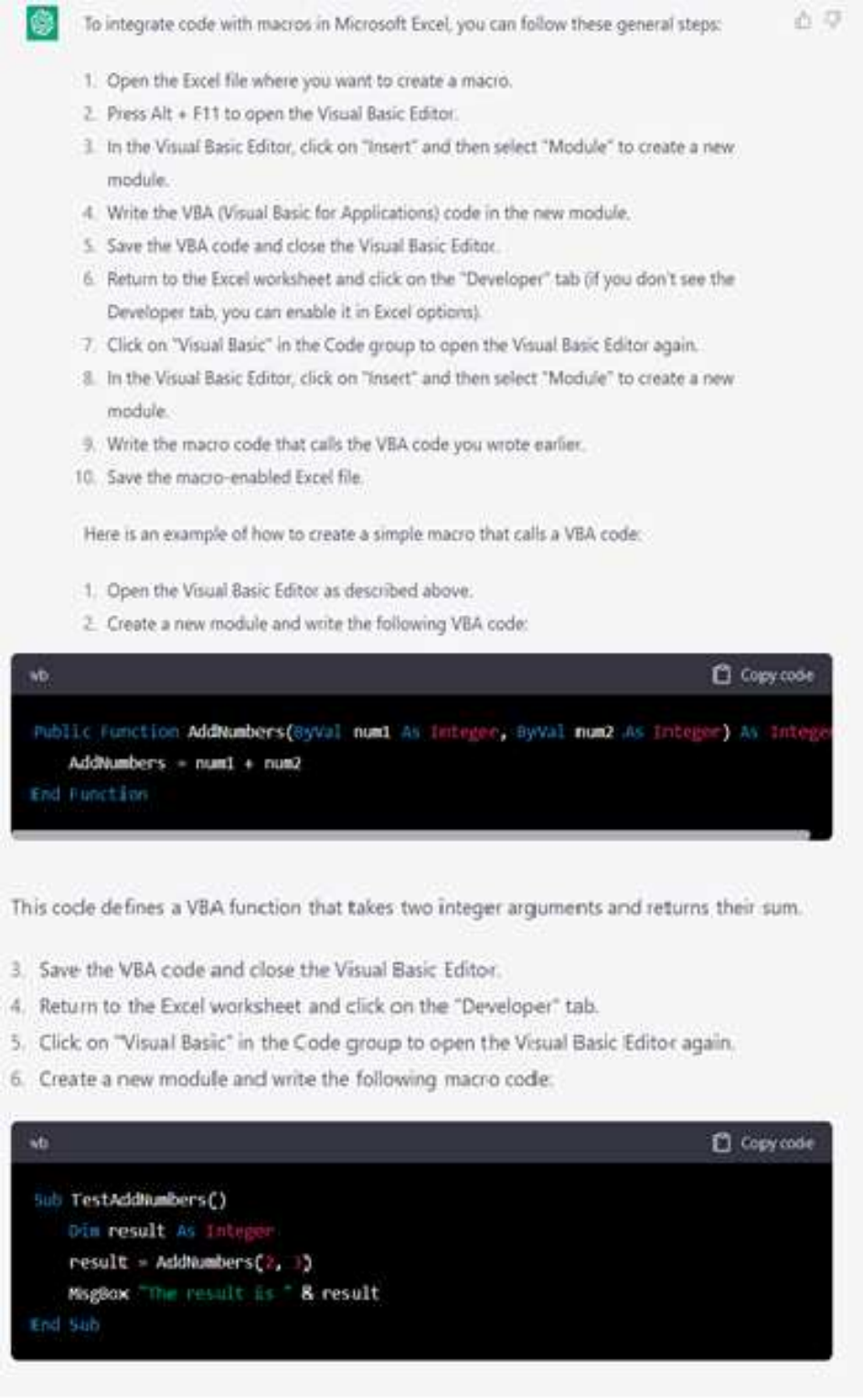

Figure 8. An example on ChatGPT integrate the code with a macro excel file

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

5.5. Vulnerability Scanning

ChatGPT can be used to scan programs with known vulnerabilities of varying complexity. C, C#, Java, and Javascript are among the languages examined. While ChatGPT is not the perfect tool for source code analysis and vulnerability detection, it is capable of understanding and generating solutions in human-like language if enough information is provided. As a result, it may be used to assist with these tasks in some capacity. ChatGPT could also be useful for developers who need to understand the purpose and function of a piece of code quickly without having to read it line by line. This could potentially solve the issue with the forbidden legacy code that lacks documentation and comments. It could be a useful debugging companion, or it could be another tool in their arsenal for discovering vulnerabilities. If ChatGPT is used for source code analysis and vulnerability detection, it is critical that the model is trained on high-quality, accurate data and that the results of its analysis are used responsibly. ChatGPT is merely a tool, and like any tool, it can be used for good or ill. It is up to the tool's users to ensure that it is used responsibly and ethically. Figure 9 depicts the use of the SAST tool for detecting vulnerabilities in Java code, with ChatGPT providing supplementary information and analyzing different scenarios to identify potential weaknesses that could be exploited by attackers. It is worth noting, however, that the purpose of this tool is not to promote or facilitate malicious activities. Instead, its goal is to enhance the security of software systems by identifying possible vulnerabilities and empowering developers to address them proactively, thereby preventing attacks from occurring in the first place. Nonetheless, attackers have been known to leverage these tips and guidelines to uncover vulnerabilities.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

Sure, here's an example Java code and how to do vulnerability detection using a SAST tool:

java Copy code

```java
import java.util.Scanner;

public class Calculator {
    public static void main(String[] args) {
        Scanner scanner = new Scanner(System.in);

        System.out.println("Enter a number: ");
        int num1 = scanner.nextInt();

        System.out.println("Enter another number: ");
        int num2 = scanner.nextInt();

        System.out.println("Choose an operation (+, -, *, /): ");
        String operation = scanner.next();

        int result = 0;

        if (operation.equals("+")) {
            result = num1 + num2;
        } else if (operation.equals("-")) {
            result = num1 - num2;
        } else if (operation.equals("*")) {
            result = num1 * num2;
        } else if (operation.equals("/")) {
            result = num1 / num2;
        }

        System.out.println("Result: " + result);
    }
}
```

Figure 9. An example of vulnerability scanning

5.6. Report Findings

ChatGPT's detection and response are impressive, demonstrating how a traditionally complex step in the attack process can now be commoditized to even the most inexperienced of hackers. Reporters use ChatGPT to automate parts of the reporting process, such as generating interview questions. This not only saves time but also allows reporters to ask more thoughtful and pertinent questions. To use ChatGPT to generate interview questions, reporters simply need to provide the model with some initial prompts and context about the interview topic. ChatGPT will then use its advanced language processing capabilities to generate a list of potential interview questions for reporters to use. Figure 10 shows an example to generate report findings by ChatGPT.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

Figure 10. An example on report findings

5.7. Vulnerability assessment in IT system

ChatGPT can be used to detect or locate code vulnerabilities. The process of identifying, quantifying, and prioritizing cybersecurity vulnerabilities in an IT system is known as vulnerability assessment. The goal of an assessment is to identify flaws that can be used to compromise systems. Errors in application code, misconfigured network devices, overly permissive access controls in a database, or arbitrary code execution, as recently detected with Log4Shell, are examples of such flaws. Vulnerability assessment is a well-established security practice. In Figure 11, ChatGPT outlines the steps for conducting a vulnerability assessment of misconfigured network devices. While these steps can be utilized by attackers to gain knowledge and carry out attacks, it's important to note that the primary purpose of vulnerability assessments is to identify and mitigate potential vulnerabilities in order to enhance the security of the system.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

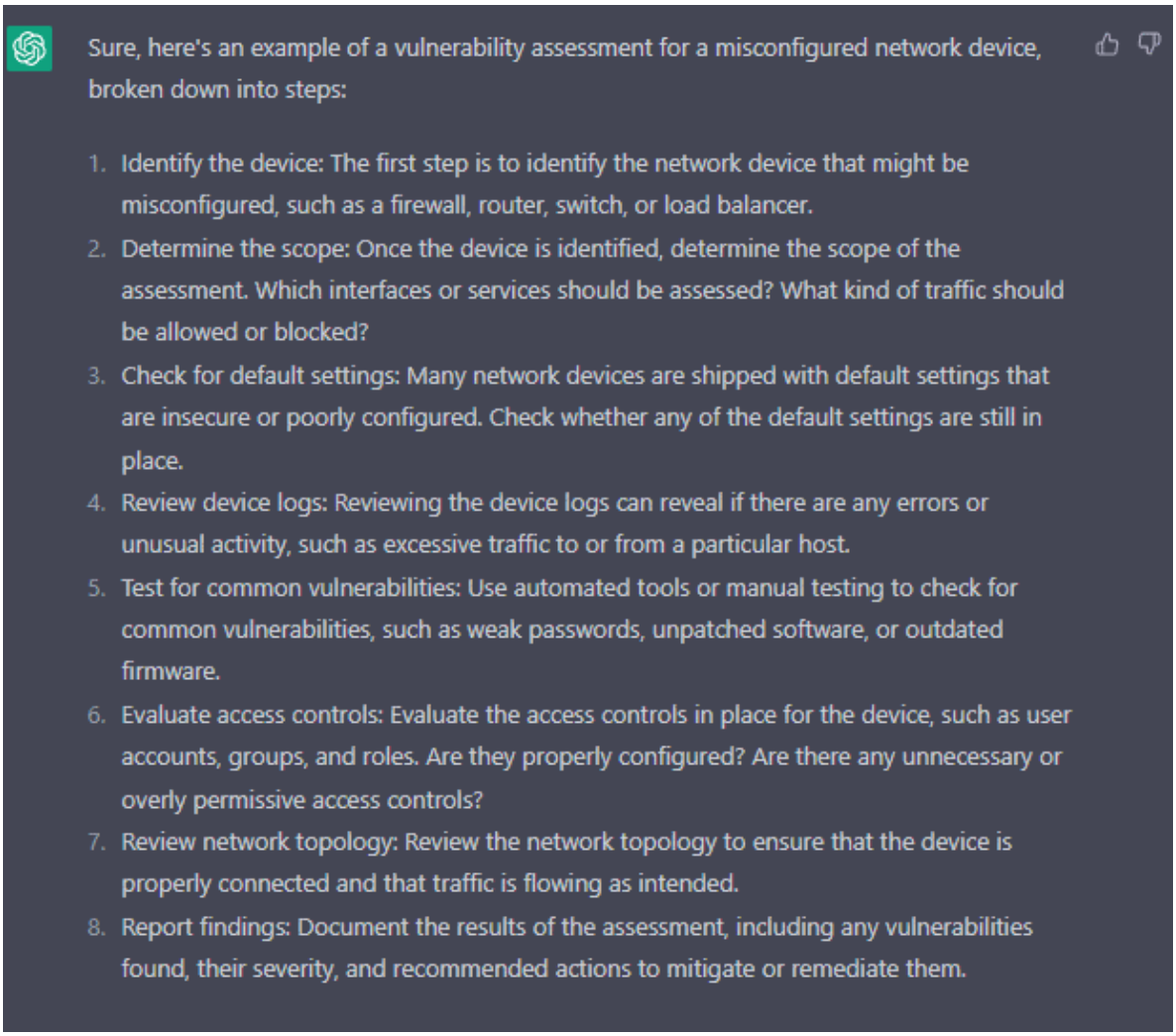

Figure 11. An example of how ChatGPT provides steps to vulnerability assessment for misconfiguration network device

5.8. Breach Notifications

An attacker can attempt to carry out a breach notification attack by sending fake breach notifications to users. The attacker may impersonate a legitimate company or website and send emails or messages to users, claiming that their account information has been compromised or that they need to reset their passwords due to a security breach. The attacker may also include a link in the message that leads to a fake login page designed to steal the user's credentials. In some cases, the attacker may even use a chatbot to send these fake breach notifications. The chatbot may be designed to imitate the company's official chatbot, making it difficult for users to distinguish between real and fake messages. Once the user provides their credentials on the fake login page, the attacker can use these credentials to gain unauthorized access to the user's account or steal their personal information.

5.9. Other ChatGPT-based Cybersecurity Issues

There are additional scenarios that can be created using ChatGPT. Here are a few examples:

- A cybersecurity issue associated with the ChatGPT model is the potential for attackers to manipulate the model's outputs in order to spread misinformation or deceive individuals. For example, an attacker could generate fake news articles or impersonate individuals through the use of the model. These attacks could have serious consequences for individuals, organizations,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

and even entire countries, making it important to develop methods to detect and prevent such attacks.

- The large size of the ChatGPT model also creates security risks in terms of data protection and privacy. The model was trained on a massive amount of diverse text data, which may include sensitive or confidential information. This creates a risk of data breaches, as well as the potential for unauthorized access to sensitive information. It is important to implement robust security measures to protect the model and the data it was trained on, as well as to prevent unauthorized access to the model.
- Businesses have become concerned about cybersecurity as the number of security breaches and ransomware attacks has increased. This has made system security more important than ever. Attackers can use ChatGPT to generate convincing phishing emails that are nearly impossible to distinguish from those sent by a real person. This can be used to trick people into clicking on malicious links or disclosing personal information.
- Cybersecurity issues [28,29] could stand as a new problem with ChatGPT for the year 2023 and beyond. Unfortunately, cybercriminals are experimenting with ChatGPT for more than just malware development. Even on New Year's Eve 2023, an underground forum member posted a thread demonstrating how they had used the tool to create scripts. The scripts could be utilized to operate an automated dark web marketplace for buying and selling stolen account details, credit card information, malware, and more. As part of the payment system for a dark web marketplace, the cybercriminal even displayed a piece of code that was produced using a third-party API. The third-party API to retrieve the most recent prices for the Monero, Bitcoin, and Ethereum cryptocurrencies.
- According to the report from a technical standpoint, "it is difficult to determine whether a specific malware was constructed using ChatGPT or not".
- Furthermore, it is difficult to determine if harmful cyber activity created with the aid of ChatGPT is currently operating in the wild. Nevertheless, as interest in ChatGPT and other AI tools grows, they will attract the attention of cybercriminals and fraudsters seeking to exploit the technology to execute low-cost and minimal-effort destructive campaigns.

ChatGPT assists security ChatGPT can assist with a variety of security-related tasks, including vulnerability management, incident response, penetration testing, and even security operations. Scientists, journalists, and simple enthusiasts find more and more uses for ChatGPT. The Awesome ChatGPT prompts website has a list of prompts (phrases to start a conversation with a bot), which can be used to "switch" ChatGPT so that it responds as Gandalf or another literary character, writes Python code, generates business letters and resumes, and even imitates a Linux terminal. Professionals in automating tedious and time-consuming tasks, optimizing security workflows, and increasing productivity. Moreover, ChatGPT keeps security professionals abreast of the most recent security trends, tools, and techniques by providing valuable insights from its extensive knowledge database. ChatGPT can be of great assistance to security professionals, allowing them to focus on more pressing tasks such as threat analysis, mitigation measures, and strategy formulation.

6. Survey Methodology and Data Collection

The survey was conducted to determine the level of ChatGPT awareness among both experts and non-experts, as well as the frequency of cybersecurity issues associated with the model's use to proffer solutions. In this study, data were collected from 5 separate domains through online questionnaires. Data collected were from different categories of experts which add up to a total sample size of 253 participants. The 5 separate domain participants include (1) public opinion, (2) knowledge management professionals, (3) information and communication technology (ICT) professionals, (4) AI and cybersecurity experts as well and (5) software engineers. The questionnaires were divided into 3 sections; Section A represents survey demographic information. Section B, information on ChatGPT capability, Section C, information on ChatGPT functionality based on features, and Section

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

D, information on ChatGPT limitations. In the design of the questionnaire, each construct carried more than 10 questions. There are 300 participants who were contacted but 253 responded to the questionnaire. Respondents were asked to indicate their agreement, disagreement, or neutrality, with several statements on the scale of measurement based on a three-point Likert scale, that is, (1) agree, (2) neutrals, and (3) disagree. The selected questions are used as mediating effects while the demographic information is used as moderating effects.

7. Results and Analysis

7.1. The background of the participants

Some of the participants were institutional leaders, IT professionals, business executives, academia, and researchers. The characteristic of the survey is presented in Table 1.

Table 1. The survey characteristics

Gender	Frequency	Percentages
Male	151	60%
Female	102	40%
Total	253	100
Participants age group	**Frequency**	**Percentages**
18-30	21	8%
31-40	43	17%
41-50	67	26%
51-55	90	36%
56 and above	32	13%
Total	253	100
Participants educational qualification	**Frequency**	**Percentages**
Undergraduate	34	9%
University Graduate	117	46%
Master's degree	80	32%
Doctorate degree	22	13%
Total	253	100
Participants' Area of Expertise	**Number of experts and non-experts in computer**	**Percentages**
Information and communication technology (ICT) professionals	45	18%
AI and cybersecurity experts	100	39%
Knowledge management professionals	35	14%
Software engineers	50	20%
Others	23	9%
Total	253	100

Table 1 revealed the survey characteristics where 60% of the participants were men and 40% were women out of 253. They are educated people from different professions with distinct age groups as shown in Table 1.

7.2. Results Based On The ChatGPT Capability

The ChatGPT application questionnaires were based on creating more awareness of its capability. Everyone user needs to be aware of ChatGPT applications and challenges. Hence, this survey questionnaire was conducted to assess participants' understanding of its applications to address human problems. The participants were non-experts in ICT, though, the questions were a bit technical,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

yet the responses were overwhelming. The results obtained from the questionnaires were presented in Table 2

Table 2. Survey Based on ChatGPT capabilities

S/n	Questions	Agreed	Neutral	Disagreed
1	ChatGPT is accessible via OpenAI' s API.	74%	8%	18%
2	ChatGPT can be used to address challenges in education and other sector.	77%	3%	23%
3	ChatGPT has many applications including teaching, learning, and researching.	72%	8%	20%
4	ChatGPT can be used only by teachers and students.	16%	5%	79%
5	ChatGPT has the capability to compose a variety of writing forms, including emails, college essays, and countless others.	87%	3%	10%
6	ChatGPT is capable of automatically composing comments for regulatory processes.	75%	6%	19%
7	It might compose letters to the editor that would be published in local newspapers.	77%	5%	18%
8	It may remark on millions of news stories, blog pieces, and social media posts per day.	81%	19%	19%
9	Capable of writing and debugging computer programs.	85%	5%	10%
10	Ability to compose poetry and song lyrics. Agreed or disagreed?	76%	4%	20%
11	Capability to write music, television scripts, fables, and student essays.	83%	7%	10%
12	Ability to compose poetry and song lyrics.	81%	9%	10%
13	Capable of remembering prior questions asked throughout the same chat.	85%	3%	12%
14	Capability to prevent offensive content from being displayed.	86%	4%	10%
15	Capable of executing moderation API and able to identify racist or sexist cues.	83%	3%	14%
16	Capable of identifying internet hate speech.	74%	5%	21%
17	Capability to construct Python-based malware that looks for popular files such as Microsoft Office documents, PDFs, and photos, then copies and uploads them to a file transfer protocol server.	79%	2%	19%
18	Ability to construct Java-based malware, which may be used with PowerShell to covertly download and execute more malware on compromised PCs.	82%	2%	16%
19	ChatGPT's knowledge of events that transpired after 2021 is limited.	72%	4%	24%
20	ChatGPT has cybersecurity challenges, that must be resolved to counter cybercriminals' exploitation.	92%	1%	7%
21	ChatGPT may generate plausible-appearing research studies for highly ranked publications even in its most basic form.	76%	2%	22%
22	Assist in software programming when fed with questions on the problem.	81%	1%	18%
23	Capability of performing coding assistance and writing job applications.	85%	5%	10%
24	Highly applicable in advertising. That is, ChatGPT has a potential use case in advertising which is in the generation of social media content.	76%	4%	20%
25	ChatGPT can be used to generate captions for video ads, which can be a powerful tool for engaging with viewers.	83%	2%	15%
26	ChatGPT can also be used in personalizing customer interactions through chatbots and voice assistants.	81%	4%	15%
27	It can be used to create more natural and personalized interactions with customers through chatbots or virtual assistants. Agreed or disagreed?	85%	5%	10%
28	ChatGPT can affect the advertising and marketing industries in a number of ways.	76%	4%	20%
29	Capability to respond to exam questions (but sometimes, depending on the test, at a level above the average human test-taker).	87%	3%	10%
30	Capacity to replace humans in democratic processes, not via voting but by lobbying.	78%	2%	20%

Based on a survey assessing the level of ChatGPT's functionality and potential use, the majority of respondents agreed with most statements, indicating that ChatGPT is accessible via OpenAI's API, can be used to address challenges in education and other sectors, has many applications including teaching, learning, and researching, and is capable of composing a variety of writing forms. However, the respondents disagreed with statements suggesting that ChatGPT can only be used by teachers and students and that it can construct Python-based malware. Some statements received mixed responses, with some agreement and some disagreement, including the statement regarding cybersecurity challenges. Overall, the analysis indicates that respondents perceive ChatGPT as a versatile and useful tool with potential applications in various contexts, but there are concerns regarding potential malicious use and cybersecurity challenges.

7.3. Results on the ChatGPT Functionality Based on Features

The questionaries were based on creating more awareness of ChatGPT functionality based on features to assess opinions on how it works. The result obtained is presented in Table 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

Table 3. Survey based on ChatGPT functionality based on features

S/n	Questions	Agreed	Neutral	Disagreed
1	ChatGPT is accessible via OpenAI' s API.	75%	6%	19%
2	ChatGPT is pre-trained on a huge corpus of conversational text, allowing it to comprehend the context of a discussion and provide responses that are more natural and coherent.	77%	5%	18%
3	ChatGPT can be used only by teachers and students.	72%	8%	20%
4	ChatGPT can handle batch input and output which allows it to handle numerous prompts and deliver several responses simultaneously.	76%	3%	21%
5	ChatGPT can be used to address challenges in education and other sector.	85%	5%	10%
6	ChatGPT has the ability to manage enormous datasets and sophisticated computations.	73%	7%	21%
7	ChatGPT is capable of generating human-like language and replies fluidly to input.	77%	5%	18%
8	ChatGPT has many applications including teaching, learning, and researching.	81%	19%	19%
9	It can track the dialogue and effortlessly handle context switching and topic shifts.	85%	3%	12%
10	It can handle both short and lengthy forms of writing.	74%	2%	24%
11	It can grasp various forms of expression including sarcasm, and irony.	81%	3 9%	10%
12	The ChatGPT model could translate text from one language to another.	83%	3%	14%
13	ChatGPT can be fine-tuned for specific conversational activities such as language comprehension.	85%	3%	12%
14	ChatGPT can be fine-tuned for text summarization, making it more successful at handling these tasks.	86%	4%	10%
15	ChatGPT can be fine-tuned for specific conversational activities such as text production.	83%	3%	14%
16	ChatGPT are a great asset to mankind because they have a lot of benefits.	74%	5%	21%
17	Although ChatGPT has many benefits, they have some barriers that limit its uses.	79%	4%	17%
18	One of the limitations of ChatGPT is the possibility of over-optimization due to its reliance on human control.	82%	2%	16%
19	ChatGPT's knowledge of events that transpired after 2021 is limited.	72%	3%	25%
20	ChatGPT has cybersecurity challenges that must be resolved to counter cybercriminals' exploitation.	82%	5%	13%

The study aimed to assess participants' opinions on ChatGPT's functionality based on its features. The results of the survey are presented in Table 4, which shows that the majority of participants agreed that ChatGPT is accessible via OpenAI's API (75%), pre-trained on a huge corpus of conversational text (77%), and can be used to address challenges in education and other sectors (85%). Additionally, most participants agreed that ChatGPT has the ability to generate human-like language and reply fluidly to input (77%), and can be fine-tuned for specific conversational activities such as language comprehension (85%) and text summarization (86%). However, the survey also revealed some limitations of ChatGPT, such as the possibility of over-optimization due to its reliance on human control (82%) and the fact that its knowledge of events that transpired after 2021 is limited (72%). Moreover, the majority of participants agreed that ChatGPT has cybersecurity challenges that must be resolved to counter cybercriminals' exploitation (82%). Overall, the study suggests that participants perceive ChatGPT as a useful tool with numerous applications. However, there is still some lack of knowledge on how it works, and cybersecurity concerns need to be addressed. Further exploration of ChatGPT by interested scholars is encouraged.

7.4. Results on the Limitations of ChatGPT

The questionnaires were based on creating more awareness of ChatGPT limitations from the diverse views of experts and users. The result obtained was presented in Table 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

Table 4. Survey based on ChatGPT limitations

S/n	Questions	Agreed	Neutral	Disagreed
1	ChatGPT is accessible via OpenAI' s API.	83%	3%	14%
2	ChatGPT has issues such as the possibility of over-optimization due to its reliance on human control.	74%	5%	21%
3	It is also constrained by a lack of knowledge regarding occurrences occurring after 2021.	79%	2%	19%
4	In other instances, it has also identified algorithmic biases in answer generation.	82%	2%	16%
5	ChatGPT is incapable of comprehending the intricacy of human language and relies solely on statistical patterns, despite its ability to produce results that appear genuine.	72%	4%	24%
6	ChatGPT has a number of deficiencies, including OpenAI admitting, however, that ChatGPT "sometimes generates plausible-sounding but inaccurate or illogical responses".	92%	1%	7%
7	The human raters are not subject matter experts; therefore, they tend to select language that appears persuasive.	76%	2%	22%
8	They could recognize many hallucinatory symptoms, but not all. Errors of fact that slip through are difficult to detect.	81%	1%	18%
9	Errors of fact that slip through are difficult to detect. In accordance with Goodhart's rule, the reward model of ChatGPT, which is based on human oversight, can be over-optimized and hamper performance.	85%	5%	10%
10	ChatGPT's knowledge of events that transpired after 2021 is limited.	75%	6%	19%
11	According to the BBC, ChatGPT will prohibit "expressing political viewpoints or engaging in political activities" as of December 2022.	77%	5%	18%
12	When ChatGPT reacts to cues containing descriptors of persons, algorithmic bias in the training data may become apparent.	72%	8%	20%
13	In one instance, ChatGPT produced a rap implying that women and scientists of color are inferior to white and male scientists, which is discriminatory.	76%	3%	21%

The analysis presented in Table 4 shows the limitations of ChatGPT from the perspective of experts and users. The survey results indicate that a significant number of participants agree with the limitations presented. The majority of respondents agree that ChatGPT has issues related to its reliance on human control, limited knowledge regarding events after 2021, algorithmic biases in answer generation, and the inability to comprehend the complexity of human language. Additionally, the survey identified limitations in the accuracy of responses, errors of fact detection, and algorithmic biases when ChatGPT reacts to cues containing descriptors of persons. Overall, the results suggest that there are several areas in which ChatGPT can be improved. The study also highlights the importance of further research on the limitations of ChatGPT to enhance its capabilities and reduce errors in response generation. Therefore, it is essential to consider these limitations while using ChatGPT to ensure that its use is appropriate and accurate.

7.5. Results On The Cybersecurity Issues Of ChatGPT

The questionnaires were based on creating more awareness of ChatGPT cybersecurity issues from the diverse views of experts and users. The result obtained was presented in Table 5

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

Table 5. Survey based on ChatGPT cybersecurity issues

S/n	Questions	Agreed	Neutral	Disagreed
1	The ChatGPT model, like any other AI system, is vulnerable to cyber security issues.	95%	2%	3%
2	ChatGPT has issues such as the possibility of being used by cybercriminals due to its reliance on human control.	84%	5%	11%
3	ChatGPT may be hacked or compromised in some way, potentially resulting in the theft or misuse of confidential material.	89%	2%	9%
4	ChatGPT poses a number of security risks. It could, for instance, be used to generate false or deceptive messages that could be used to deceive people or organizations.	92%	2%	6%
5	One of the main concerns about using the ChatGPT model is the potential for malicious use, such as creating fake news, impersonating individuals, or spreading disinformation.	82%	4%	14%
6	The model may have access to sensitive data that, if it falls into the wrong hands, could be misused.	94%	1%	5%
7	The ChatGPT model's large size also poses security risks in terms of data protection and privacy.	86%	2%	12%
8	Attackers can use ChatGPT to create convincing phishing emails that are nearly impossible to distinguish from those sent by a real person.	81%	1%	18%
9	Cyberattackers are already planning on how to use ChatGPT for malware development, social engineering, disinformation, phishing, malvertising, and money-making schemes.	85%	5%	10%
10	As interest in ChatGPT and other AI tools grows, so will cybercriminals and fraudsters looking to use the technology to carry out low-cost, low-effort destructive campaigns.	75%	6%	19%
11	It creates a risk of data breaches, as well as the potential for unauthorized access to sensitive information.	77%	5%	18%
12	Cyberattacks can have serious consequences for individuals, organizations, and even entire countries, so developing methods to detect and prevent such attacks is critical.	84%	6%	10%
13	Companies need to be mindful of the numerous risks associated with ChatGPT and take precautions to mitigate them.	87%	5%	8%
14	Companies should invest in robust cybersecurity measures and keep up with security trends.	86%	3%	11%
15	Organizations can reap the many benefits of ChatGPT while defending themselves against those who abuse the tool by implementing appropriate safeguards.	96%	1%	3%
16	It is important to implement robust security measures to protect the model and the data it was trained on, as well as to prevent unauthorized access to the model.	89%	4%	7%
17	Spam detection and anti-phishing tools can be used to scan the text of an email for common indicators of fraudulent behavior regularly.	93%	2%	5%
18	In the new world of ChatGPT, penetration testing can help to safeguard data and ensure its confidentiality, integrity, and availability.	81%	5%	14%
19	Businesses must also improve their data resilience strategy and implement a solid data protection plan.	88%	3%	9%
20	A major focus for the future of ChatGPT should be the development of methods to detect and prevent malicious use of the model.	87%	5%	8%

The results presented in Table 5 indicate that there is widespread concern about cybersecurity issues associated with the ChatGPT model. The survey was designed to create more awareness of ChatGPT cybersecurity issues from the diverse views of experts and users. The responses show that a vast majority of the participants agreed that ChatGPT is vulnerable to cybersecurity issues (95%). This suggests that there is a general understanding that AI systems like ChatGPT can be exploited by cybercriminals. Another issue that the respondents expressed concern about was the potential for malicious use of ChatGPT, such as creating fake news, impersonating individuals, or spreading disinformation. This was agreed by 82% of the participants. Additionally, respondents were also worried about the possibility of attackers using ChatGPT to create convincing phishing emails that are nearly impossible to distinguish from those sent by a real person (81%). Furthermore, the respondents indicated that ChatGPT poses several security risks, which include the possibility of it being hacked or compromised, resulting in the theft or misuse of confidential material (89%). There was also a concern about the model having access to sensitive data that, if it falls into the wrong hands, could be misused (94%). It is clear from the results that the use of ChatGPT poses a significant threat to data protection and privacy (86%). The large size of the model was also identified as a potential risk. Respondents expressed concern about the potential for cyber attackers to use ChatGPT for malware development, social engineering, disinformation, phishing, malvertising, and money-making schemes (85%). The survey results also suggest that organizations need to be mindful of the numerous risks associated with ChatGPT and take precautions to mitigate them (87%). This includes investing in robust cybersecurity measures and keeping up with security trends (86%). Additionally, it is important to implement robust security measures to protect the model and the data it was trained on, as well as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

to prevent unauthorized access to the model (89%). Spam detection and anti-phishing tools can be used to scan the text of an email for common indicators of fraudulent behavior regularly (93%). Finally, the results of the survey indicate that there is widespread concern about the cybersecurity issues associated with the use of ChatGPT. It is important for organizations to take appropriate measures to protect themselves from the risks associated with the use of this technology. Further research is required to develop methods to detect and prevent the malicious use of ChatGPT and to improve the overall cybersecurity of AI systems.

7.6. Discussion

The data presented in Table 1 shows that both males and females participated in the survey, with a higher percentage of males compared to their female counterparts. However, it is important to note that the study was an open online survey, indicating that there was no discrimination based on gender. The survey participants were drawn from various careers and age groups, providing a diverse sample for the study.

Table 2 indicates that most of the respondents were not aware of ChatGPT's capabilities, benefits, and limitations. This highlights the need for more awareness and education for the general public, including institutions, teachers, learners, and researchers, on the current development in the ICT industry. It is important to raise awareness of the potential uses and limitations of ChatGPT, which could help individuals and organizations make informed decisions on how to leverage the technology.

Table 3 presents evidence that users are already seeing a positive return on their investment in ChatGPT, despite the chatbot still being in its early stages of development. However, the findings also highlight the importance of investigating the differences in understanding how the chatbot operates, particularly with regard to cybersecurity issues. Most participants acknowledged that ChatGPT faces cybersecurity challenges and these issues must be addressed to prevent exploitation by cybercriminals.

Table 4 demonstrates that while ChatGPT is capable of generating creative content for ads and marketing campaigns, there are limits to its capabilities that may require further research. It is also worth noting that the impact of ChatGPT on the industry will depend on how it is implemented and used by businesses.

Table 5 provides insight into the cybersecurity issues associated with ChatGPT from the perspectives of both experts and users. The data highlights that cybersecurity is a major concern with ChatGPT and that an immediate solution is required. The study also revealed that there is a high level of awareness of ChatGPT among computer experts, but this is not necessarily the case for non-experts. Therefore, there is a need for ongoing education and awareness-raising efforts to address cybersecurity issues associated with the use of ChatGPT.

Overall, the findings from the study suggest that ChatGPT is a unique and powerful tool that has the potential to revolutionize the way we handle human problems on the internet. However, it is important to be aware of the limitations and cybersecurity issues associated with its use. By understanding these issues, we can ensure that ChatGPT is used in a safe and responsible manner that benefits society as a whole.

8. Recommendations and Solutions

In this section, we offer recommendations for enterprises considering the integration of ChatGPT into their services. Furthermore, we present solutions to prevent potential cyberattacks that may arise from the misuse of ChatGPT.

8.1. Recommendations

We recommend that enterprises and information and communication technology (ICT) or information and technology (IT) departments contemplating the addition of ChatGPT and other generative AI technologies to their service catalogs adhere to four essential suggestions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

- To restrict its abuse, we can only hope that authorities will move swiftly and that tools and laws will follow at the same rate as the underlying technology.
- ChatGPT and other generative AI solutions are tools, and like any piece of software or computer hardware, this technology excels at certain tasks while failing in others. It is crucial for IT leaders to comprehend the business processes and the areas in which ChatGPT excels so that they can identify opportunities to minimize friction, improve the quality of the service experience, and increase productivity.
- ICT teams should begin with augmentation, despite the allure of jumping directly into the AI endgame. The great technology that is generative AI is not self-sufficient. It still needs supervision and feedback from human curators, and as more processes are enhanced, IT should anticipate incurring the cost of enhancing its supervisory skills.
- New frontiers bring with them new potential issues. Workflow-managing chatbots are not complicated. However, a bot that interacts with users and generates content could subject the organization to legal risk by, for example, duplicating content from other sources or producing obscene or offensive photos. Before deploying such technology, organizations should get legal guidance.
- The capabilities of generative AI such as ChatGPT can significantly minimize the amount of manual labor required to complete specific jobs. Nevertheless, any function or task that necessitates broad permissions or approval and highly specialized or contextual knowledge may expose an organization to risk.
- Artificial intelligence can improve teaching and learning, but it should not be used as a replacement.
- Literary works should be cited and properly referenced; the content provided by ChatGPT is the work of authors/researchers/websites and is not created out of thin air.
- Proper research should be conducted rather than relying solely on the responses provided by ChatGPT.
- Despite the limited number of responses provided by ChatGPT, search engines remain a reliable source of information and should not be substituted/replaced.

8.2. Solution to the use of ChatGPT for Cyberattack

We must emphasize that every organization or company employee must contribute to the security of their infrastructure. We recommend that organizations double down on best practices such as the following:

- Train employees as soon as possible on new threats, including fun, in-person sessions that help make security habits relatable.
- Strengthen their security by discussing cybercrime and its potential impact on an organization's ability to operate on a regular basis.
- Consider adding security services to contain threats and monitor for potential issues that get past defend-and-protect solutions, such as stopping infiltrations that can occur because of phishing attacks.
- Penetration testing can help you safeguard your data and ensure its confidentiality, integrity, and availability in the new world of ChatGPT. Businesses must also strengthen their data resilience strategy and have a solid data protection plan in place.
- While, ChatGPT has many advantages for businesses, it also has significant security risks. Organizations must be aware of these risks and take precautions to reduce them. They should invest in strong cybersecurity measures and stay current on security trends. By implementing appropriate safeguards, organizations can reap the many benefits of ChatGPT while defending themselves against those who misuse the tool.
- Spam detection and anti-phishing tools routinely scan the text of an email for common indicators of fraudulent behavior, such as spelling and grammatical errors, particularly those produced by people with limited English proficiency. These filters are ineffective against phishing emails generated by ChatGPT because the text generated by it does not contain these errors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

9. Conclusion

In conclusion, ChatGPT is an impressive model that has the potential to address a wide range of human needs such as text writing, learning, and researching. However, it is important to note that its widespread use has also led to concerns about its malicious use in areas such as IT system vulnerabilities, malware attacks, SQL injection, social engineering/phishing, macros & LOLBIN, breach notifications, and more. As the survey findings showed, there is a high level of awareness of ChatGPT among computer experts, but the level of awareness among non-experts is still low. Furthermore, the survey revealed that cybersecurity complications associated with the model's applications cannot be ignored, and organizations must play a significant role in protecting their computer infrastructure. While ChatGPT is a significant step in the progress of artificial intelligence, its generative skills have also raised concerns about its potential misuse, particularly from an Intellectual Property standpoint. It is essential to recognize that ChatGPT is an incremental step in NLP research, and businesses will gradually reap the benefits through advancements to fundamental NLP tasks. However, most businesses lack the knowledge and infrastructure required to use ChatGPT's main technology components, which may require additional work in the form of mature tools and APIs to facilitate the process of fine-tuning pre-trained LLMs to enterprise-specific data and domains. In summary, the future of ChatGPT looks promising, but it is crucial to be aware of the cybersecurity challenges associated with its applications and the potential for malicious use. Organizations and users must remain vigilant and take necessary precautions to protect their computer infrastructure.

Author Contributions: Conceptualization, methodology, formal analysis M.A, O.I.A and B.A; writing—review and editing, writing—original draft preparation O.I.A, M.A, and A.M; supervision, project administration O.I.A and M.A; data curation, validation software A.M and R.A; funding acquisition M.A and B.A.; investigation, resources M.A, A.A and O.E.A. All authors have read and agreed to the published version of the manuscript

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This article contains no data or material other than the articles used for the review and referenced.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Schulman, J.; Zoph, B.; Kim, C.; Hilton, J.; Menick, J.; Weng, J.; Uribe, J.; Fedus, L.; Metz, L.; Pokorny, M.; others. ChatGPT: Optimizing language models for dialogue, 2022.
2. Sobania, D.; Briesch, M.; Hanna, C.; Petke, J. An Analysis of the Automatic Bug Fixing Performance of ChatGPT. *arXiv preprint arXiv:2301.08653* **2023**.
3. Ventayen, R.J.M. OpenAI ChatGPT Generated Results: Similarity Index of Artificial Intelligence-Based Contents. *Available at SSRN 4332664* **2023**.
4. Frieder, S.; Pinchetti, L.; Griffiths, R.R.; Salvatori, T.; Lukasiewicz, T.; Petersen, P.C.; Chevalier, A.; Berner, J. Mathematical Capabilities of ChatGPT. *arXiv preprint arXiv:2301.13867* **2023**.
5. Qadir, J. Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education **2022**.
6. Jiao, W.; Wang, W.; Huang, J.t.; Wang, X.; Tu, Z. Is ChatGPT a good translator? A preliminary study. *arXiv preprint arXiv:2301.08745* **2023**.
7. Black, S.; Biderman, S.; Hallahan, E.; Anthony, Q.; Gao, L.; Golding, L.; He, H.; Leahy, C.; McDonell, K.; Phang, J.; others. Gpt-neox-20b: An open-source autoregressive language model. *arXiv preprint arXiv:2204.06745* **2022**.
8. Dahiya, M. A Tool of Conversation: Chatbot, International Journal of Computer Sciences and Engineering, Volume-5, Issue-5 E-ISSN: 2347-2693. *Int. J. Comput. Sci. Eng.* **2017**, *5*.
9. Vaishya, R.; Misra, A.; Vaish, A. ChatGPT: Is this version good for healthcare and research? *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **2023**, *17*, 102744. doi:https://doi.org/10.1016/j.dsx.2023.102744.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 September 2023 doi:10.20944/preprints202309.1768.v1

10. Sallam, M. The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. *medRxiv* **2023**, pp. 2023–02.
11. Biswas, S.S. Role of chat gpt in public health. *Annals of biomedical engineering* **2023**, *51*, 868–869.
12. DEMİR, Ş.Ş.; DEMİR, M. Professionals' perspectives on ChatGPT in the tourism industry: Does it inspire awe or concern? *Journal of Tourism Theory and Research* **2023**, *9*, 61–76.
13. Eke, D.O. ChatGPT and the rise of generative AI: threat to academic integrity? *Journal of Responsible Technology* **2023**, *13*, 100060.
14. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; others. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* **2023**, *103*, 102274.
15. Hill-Yardin, E.L.; Hutchinson, M.R.; Laycock, R.; Spencer, S.J. A Chat (GPT) about the future of scientific publishing. *Brain Behav Immun* **2023**, *110*, 152–154.
16. Surameery, N.M.S.; Shakor, M.Y. Use chat gpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering (IJITC) ISSN: 2455-5290* **2023**, *3*, 17–22.
17. Eggmann, F.; Weiger, R.; Zitzmann, N.U.; Blatz, M.B. Implications of large language models such as ChatGPT for dental medicine. *Journal of Esthetic and Restorative Dentistry* **2023**.
18. Biswas, S.S. Potential use of chat gpt in global warming. *Annals of biomedical engineering* **2023**, *51*, 1126–1127.
19. Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepaño, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G.; Maningo, J.; others. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digital health* **2023**, *2*, e0000198.
20. Qi, Y.; Zhao, X.; Huang, X. safety analysis in the era of large language models: a case study of STPA using ChatGPT. *arXiv preprint arXiv:2304.01246* **2023**.
21. Ferrara, E. Should chatgpt be biased? challenges and risks of bias in large language models. *arXiv preprint arXiv:2304.03738* **2023**.
22. Hosseini, M.; Horbach, S.P. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other Large Language Models in scholarly peer review. *Research Integrity and Peer Review* **2023**, *8*, 4.
23. Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* **2023**, *3*, 121–154. doi:https://doi.org/10.1016/j.iotcps.2023.04.003.
24. Bhattaram, S.; Shinde, V.S.; Khumujam, P.P. ChatGPT: The next-gen tool for triaging? *The American Journal of Emergency Medicine* **2023**, *69*, 215–217. doi:https://doi.org/10.1016/j.ajem.2023.03.027.
25. Wu, T.; He, S.; Liu, J.; Sun, S.; Liu, K.; Han, Q.L.; Tang, Y. A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA Journal of Automatica Sinica* **2023**, *10*, 1122–1136. doi:10.1109/JAS.2023.123618.
26. Van Dis, E.A.; Bollen, J.; Zuidema, W.; van Rooij, R.; Bockting, C.L. ChatGPT: five priorities for research. *Nature* **2023**, *614*, 224–226.
27. Taofeek, O.T.; Alawida, M.; Alabdulatif, A.; Omolara, A.E.; Abiodun, O.I. A Cognitive Deception Model for Generating Fake Documents to Curb Data Exfiltration in Networks During Cyber-Attacks. *IEEE Access* **2022**, *10*, 41457–41476. doi:10.1109/ACCESS.2022.3166628.
28. Alawida, M.; Omolara, A.E.; Abiodun, O.I.; Al-Rajab, M. A deeper look into cybersecurity issues in the wake of Covid-19: A survey. *Journal of King Saud University-Computer and Information Sciences* **2022**.
29. Abiodun, O.I.; Alawida, M.; Omolara, A.E.; Alabdulatif, A. Data provenance for cloud forensic investigations, security, challenges, solutions and future perspectives: A survey. *Journal of King Saud University-Computer and Information Sciences* **2022**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.