







CHATGPT: A DOUBLE-EDGED SWORD IN CYBERSECURITY - EVALUATING RISKS AND RECOMMENDATIONS FOR SAFER AI INTEGRATION

Dr. Mitesh G Patel¹, Dr. Hinal N Prajapati², Nihar K Patel³, Nirmal S Patel⁴, Anand K Patel⁵, Hemali A Brahmabhatt⁶

¹ Assistant Professor, Asian BCA College, Vadali

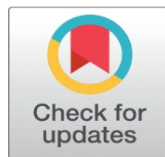
² Assistant Professor, Department of Computer & IT, HNGU-Patan

³ Assistant Professor, Asian BCA College, Vadali

⁴ I/c Principal, Asian Institute of Technology, Vadali

⁵ Head of Department, Asian Institute of Technology, Vadali

⁶ Assistant Professor, M.L Gandhi BCA College, Modasa



Corresponding Author

Dr. Mitesh G Patel,

mca.mitesh@gmail.com

DOI

[10.29121/shodhkosh.v5.i5.2024.1956](https://doi.org/10.29121/shodhkosh.v5.i5.2024.1956)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2024 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.



ABSTRACT

Over the years, natural language processing (NLP) has seen remarkable progress, largely thanks to the advancements in artificial intelligence (AI). Specifically, recent strides in this field can be attributed to the emergence of sophisticated conversational AI systems like ChatGPT. Since its release in November 2022, ChatGPT has captivated millions of users with its impressive features and capabilities. However, there's a growing concern about its potential misuse by malicious actors. In particular, ChatGPT opens up new avenues for hackers to compromise cybersecurity. This article delves into a comprehensive exploration of how ChatGPT can significantly aid hackers in executing various attacks. The investigation draws from cutting-edge research in this domain. Additionally, we evaluate ChatGPT's impact on cybersecurity, both positive and negative. The conclusion is clear: ChatGPT has indeed facilitated hacking behaviors and could be exploited for malicious purposes. To mitigate these risks, continuous development and the establishment of appropriate standards are crucial. Policymakers and developers must collaborate, taking into account user concerns and the responsible use of this powerful tool. Ultimately, this research article offers insightful discussions and recommendations to enhance AI-based systems.

Keywords: ChatGPT, Chatbot, Cybersecurity, Security Attacks, Security Threats, Phishing Attack, Social Engineering, Malicious Activities, Malware.

1. INTRODUCTION

ChatGPT, a cutting-edge linguistic model developed by OpenAI, is among the most fascinating advancements in the field of artificial intelligence (AI). ChatGPT has already made a significant mark and is anticipated to expand rapidly in the forthcoming years due to its ability to generate text that mimics human conversation and responds to complex inquiries. The potential of ChatGPT and extensive language models to enhance our lives changes the way we interact with

technology (Aljanabi, 2023). ChatGPT is a conversational AI that was launched in November 2022. It is trained on a vast amount of data to be proficient in responding to questions (Gill & Kaur, 2023; Sebastian, 2023b).

ChatGPT is a system that amalgamates extensive natural language processing (NLP) algorithms and artificial intelligence (AI) to offer an engaging dialogue interface. This allows users to input in everyday language and receive comprehensible replies. Within this framework, ChatGPT assesses the user's input and generates an appropriate response (Sharma & Dash, 2023). ChatGPT represents a remarkable advancement in the realm of artificial intelligence (Roumeliotis & Tselikas, 2023) and has evolved into a powerful instrument with a broad range of uses across multiple domains (Ray, 2023) such as (Ali & Aysan, 2023; Biswas, 2023a, 2023b; Choi, Hickman, Monahan, & Schwarcz, 2023; Kshetri, 2023; Sallam; Shoufan, 2023; Sun & Yao, 2023; Surameery & Shakor, 2023; Xames & Shefa, 2023).

Thanks to its exceptional capacity to generate believable and consistent replies to a wide range of subjects, ChatGPT has gained global fame, acknowledgement, and interest (Khosravi, Shafie, Hajiabadi, Raihan, & Ahmed, 2023). Depending on one's perspective, ChatGPT can be seen as a remarkable leap forward for mankind or a serious peril. However, it is primarily viewed in the realm of cybersecurity as a hazard that enables novices to become skilled hackers (Mansfield-Devine, 2023; Marshall, 2023). With just a few clicks and minutes, one can acquire all the necessary tools for phishing or other forms of attack (Grbic & Dujlovic, 2023). ChatGPT poses substantial cybersecurity threats that must be tackled (Addington, 2023). In the work of Sebastian (2023a), it was pointed out that as AI Chatbots and similar technologies gain popularity, the potential security risks and related cyber threats increase. There's a danger that ChatGPT could provide cybercriminals with straightforward access to scripting and coding, thereby lowering the barriers to entry in this domain. Despite the existence of safeguards to prevent harmful users from obtaining these scripts and code, it remains crucial to consistently assess and keep an eye on the risks they present and the countermeasures in place, given the constant evolution of technology.

This research seeks to explore the cutting-edge potential of misusing ChatGPT for cyber-attacks. The aim is to highlight the potential abuse of this tool by malicious entities. Numerous recent publications have underscored the positive impact of ChatGPT across various domains, emphasizing its ability to expedite the processing and manipulation of vast data sets, yielding exceptional outcomes. Yet, within the realm of cybersecurity, it's often portrayed as a method for enhancing security breaches (Thorncharoensri, P., 2019). This tool can swiftly enable novice hackers to gain advanced skills. The importance of this research lies in assisting ChatGPT developers and cybersecurity experts in implementing necessary measures that could limit the malicious use of this tool, while enhancing its functionalities to bolster cybersecurity. ChatGPT, being in its nascent stages of development, could escalate cybersecurity risks if not adequately safeguarded. The research scrutinizes various facets of cybersecurity, including technical vulnerabilities, social engineering, malware threats, phishing attacks, identity theft, and other cybersecurity dimensions.

This segment presents an overview of ChatGPT, a remarkable innovation in the realm of AI that has gained considerable traction and amassed a significant user base globally. The structure of this document is as follows. Section 2 offers a succinct depiction of ChatGPT's capacity to supplant human labor. Section 3 explores the role of ChatGPT in bolstering cybersecurity, considering that AI systems contribute substantially to enhancing cybersecurity through the automation of threat detection and response, comprehensive data scrutiny, and the provision of robust tools for predicting attacks. Section 4 delves into the potential hazards of ChatGPT as an instrument that could facilitate the generation of diverse security breaches, thereby posing a threat to the field of cybersecurity. This section conducts an exhaustive examination of the possible attacks that could leverage ChatGPT and outlines the prevalent attacks in this scenario. Furthermore, this section highlights the imperative to refine ChatGPT by rectifying its weaknesses, which are manipulated by malicious entities, and enforcing suitable and efficacious regulations. A detailed discourse and suggestions are put forth in Section 5. Lastly, the conclusions are drawn in Section 6.

2. INFLUENCE OF CHATGPT ON WORKPLACE DYNAMICS

While generative AI can enhance efficiency, it also poses a threat to human employment. Its most revolutionary effects have been seen in scholarly research, education, and learning. Moreover, Dwivedi et al. (2023) highlighted that issues such as biases, obsolete training data, lack of trustworthiness, and absence of transparency are significant worries. It's vital to select and enforce regulations to avoid the abuse and manipulation of generative AI. George and George (2023) stated that a large number of specialists agree that substantial technological progress is needed before machines can fully take over all human jobs.

3. CAN CHATGPT ENSURE SECURITY AND CONTRIBUTE TO ENHANCED CYBERSECURITY?

Broadly speaking, ChatGPT is a double-edged sword in the realm of security. It's viewed by some scholars as a potent instrument for enhancing cybersecurity (Temara, 2023). Derner and Batistič (2023) pointed out that ChatGPT employs methods to progressively increase its reliability and security. These methods encompass a comprehensive strategy to tackle issues arising from antagonistic behavior, RLHF that facilitates model enhancement over time, and a thorough fine-tuning process on a handpicked dataset that aids in restricting its outputs to pertinent and secure content. Notwithstanding these precautions, there exist methods to bypass ChatGPT's safeguards, for instance, through inventive instruction. The filters of ChatGPT are engineered to prevent the production of damaging or inappropriate outputs (Derner & Batistič, 2023).

In order to gather ChatGPT's perspective on its contribution to enhancing cybersecurity a questions outlined in Table 1 was posed to ChatGPT.

Table 1: Examining ChatGPT's Response to Inquiries About its Role in Cybersecurity Advancements

Questions	ChatGPT Response
Can ChatGPT play a significant role in improving online security measures?	Certainly! ChatGPT has the potential to play a significant role in improving online security measures. Its capabilities in natural language processing (NLP) can be harnessed for various security applications such as: Threat Detection, Vulnerability Assessment, Security Training and Education, Automated Response, Policy Enforcement. Overall, ChatGPT's natural language processing capabilities make it a valuable tool for enhancing cybersecurity measures by assisting in threat detection, vulnerability assessment, security training, automated response, and policy enforcement.
Does the implementation of ChatGPT offer any advantages for cyber threat detection and prevention?	Yes, the implementation of ChatGPT offers several advantages for cyber threat detection and prevention like Natural Language Understanding, Scalability, Anomaly Detection, Adaptability, Integration. Overall, the implementation of ChatGPT offers significant advantages for cyber threat detection and prevention by leveraging its natural language understanding, scalability, anomaly detection capabilities, adaptability, and integration potential.
Does ChatGPT have a positive contribution to enhance cyber security?	Yes, ChatGPT can make a positive contribution to enhancing cybersecurity in several ways: Threat Detection, Anomaly Detection, Vulnerability Assessment, Security Training, Automated Response. Overall, ChatGPT's natural language processing capabilities make it a valuable tool for enhancing cybersecurity by assisting in threat detection, anomaly detection, vulnerability assessment, security training, and automated response. However, it's important to note that while ChatGPT can be a useful component of a comprehensive cybersecurity strategy, it should be used in conjunction with other security measures and best practices for optimal effectiveness.

The feedback from ChatGPT suggested that ChatGPT could potentially be employed to orchestrate more intricate assaults. Nonetheless, it can also serve as a resource for investigators to gain a deeper understanding and foresee such attacks, thereby paving the way for more robust cybersecurity strategies.

4. REVIEW OF EXISTING LITERATURE

ChatGPT has garnered considerable attention since its launch in November 2022 due to its ability to generate conversational replies. This highlights the capabilities of AI systems, but it also brings several risks to the forefront. This part of the discussion explores the potential of ChatGPT as a tool that could facilitate various security breaches, posing a significant threat to cybersecurity. As ChatGPT was introduced in late 2022, the majority of the references in this paper are current and represent the cutting-edge in this domain. Large language models could potentially be misused to assist in the creation of malware and the drafting of phishing emails. As a result, entities involved in the development of large language models must exert considerable effort to reduce the likelihood of their misuse. Moreover, enhancing anti-malware measures is crucial to protect patient information, as well as the operational software and hardware utilized by healthcare institutions, considering that malicious actors are adept at exploiting technologies like large language models (Eggmann, Weiger, Zitzmann, & Blatz, 2023).

ChatGPT can be a valuable tool in enhancing cybersecurity, identifying threats, monitoring security, educating about security, and analyzing malware. Yet, it's important to be mindful of potential issues such as context insensitivity, excessive dependence on technology, security risks, restricted functionalities, and inherent biases (Biswas, 2023c).

Yang et al. (2023) pointed out that chatbots are incapable of learning from encrypted data. As a result, the act of decrypting data for training could potentially expose information to those it was not intended for. Moreover, a tailored strategy is necessary to tackle the unique security challenges presented by different chatbot contexts and situations. For example, chatbots deployed in healthcare settings might necessitate distinct security measures compared to those employed in the financial industry (Yang, Chen, Por, & Ku, 2023).

Machine-learning frameworks are vulnerable to various attacks (Elsadig, 2023; Elsadig & Gafar, 2022), such as membership inference, model theft, model tampering, and input distortion. As a result, there's a need for more resilient adversarial machine-learning systems. These adversarial systems evaluate attack strategies and defensive mechanisms to prevent the misuse of these systems for malicious purposes. In the context of ChatGPT, an adversarial input prompt could compel the model to generate harmful text or damaging sentences that present substantial security risks or disseminate false information (B. Liu et al., 2023).

Through the automation of routine operation tasks, ChatGPT can lead to significant time and cost savings. Nevertheless, ChatGPT could potentially yield prejudiced and deceptive results, provoke ethical dilemmas, and be misused. Therefore, it is crucial to recognize these hazards and implement protective measures. As a result of these concerns, certain nations have banned the utilization of ChatGPT (Bahrini et al., 2023).

ChatGPT, in addition to interpreting text in a human-like way, has the ability to translate natural language into programming code. This essentially means it assists in generating code (C. Liu et al., 2023; Rahman & Watanobe, 2023). Nonetheless, it's crucial to evaluate the security of the applications developed by ChatGPT. In order to examine the safety of the generated code, a variety of software were developed using ChatGPT (Khoury, Avila, Brunelle, & Camara, 2023). It was highlighted that despite being cognizant of potential weaknesses, ChatGPT often generates code that is susceptible to breaches.

Phishing involves masquerading as a trustworthy entity to pilfer confidential information like usernames, passwords, and credit card details, typically with the aim of inflicting harm on an individual or organization. Indeed, crafting phishing attacks is challenging without technical know-how. Yet, with ChatGPT, this can be accomplished by posing a few straightforward queries to the bot. Grbic and Dujlovic (2023) demonstrated a practical instance of employing ChatGPT to construct a counterfeit login page resembling a Facebook login page, which was then used to acquire Facebook credentials. This instance underscores the ease with which the login page of any application can be replicated to steal the login data that grants access to the targeted application. Charan et al. noted that ChatGPT equips attackers with the capability to initiate complex and targeted attacks more rapidly. Moreover, this technology offers novice attackers extra tools to execute a range of attacks and motivates script kiddies to develop tools that can expedite the expansion of cybercrime (Charan, Chunduri, Anand, & Shukla, 2023).

A honeypot is a crucial instrument in cyber security, utilized to detect, halt, and scrutinize malicious and detrimental activities within a computer network. Essentially, it's a snare designed to lure potential attackers, whose actions are then observed and recorded for subsequent threat analysis. In this scenario, ChatGPT offers an innovative tool that can serve as a potential honeypot interface in cyber security. It's possible to create a dynamic environment that can adapt to the maneuvers of attackers and shed light on their strategies, methods, and procedures by emulating Linux, Mac, and Windows terminal commands and providing an interface for TeamViewer, nmap, and ping (McKee & Noever, 2023).

However, ChatGPT's ability to mimic these terminals and other application interfaces empowers hackers to execute advanced attacks.

Sebastian (2023a) carried out a study on specific facets of ChatGPT security concerns. The results they reported suggested that social engineering attacks were identified as the foremost chatbot cyber risks, succeeded by malware threats. These findings are enumerated in below Chart 1.

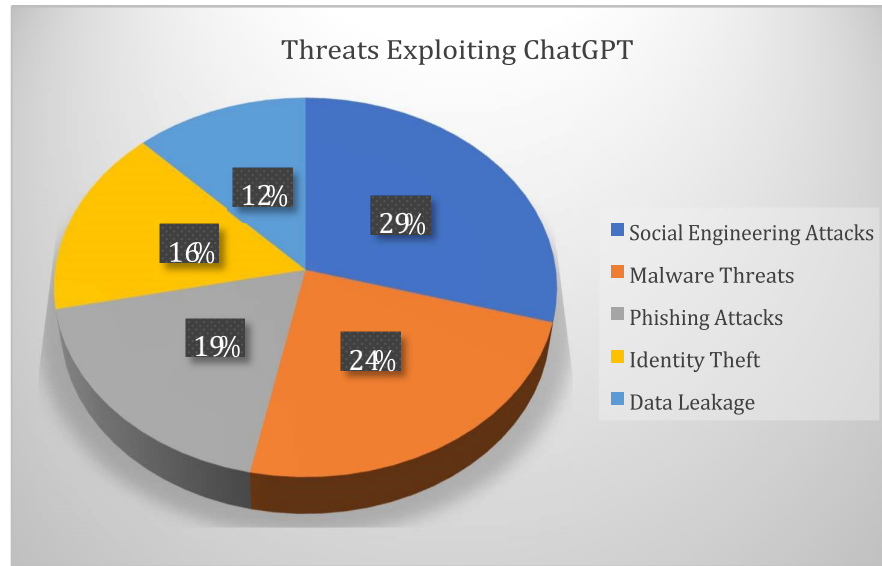


Chart 1 : Threats Exploiting ChatGPT

Sebastian pointed out that the privacy threat posed by large language models like ChatGPT is a multifarious and intricate issue. Hence, a blend of various methods, along with continuous research and development, is required to mitigate these risks (Sebastian, 2023b). They evaluated privacy-enhancing technologies (PETs) and carried out a survey to gauge the apprehensions of chatbot users regarding data privacy when utilizing applications built on large language models. Their findings underscore the urgent necessity for collaborative initiatives to bolster data security and privacy in AI systems. This research underscores the importance of continuous exploration, study, regulation, and application of PETs in AI models (Sebastian, 2023b).

While Sharma and Dash (2023) acknowledged some advantages of ChatGPT, they also highlighted that ChatGPT presents considerable cybersecurity risks and is a magnet for cybercriminals. It could potentially evolve into a conduit for attackers to swiftly initiate cyberattacks. The authors provided several instances demonstrating that ChatGPT can aid cybercriminals in executing harmful actions. These instances include Infostealer and Encryption Tools. The former is a malware based on Python, and the latter is a Python script that carries out encryption and decryption. Both were developed using ChatGPT. Furthermore, the authors noted that ChatGPT aids in committing fraud. They concluded that despite ChatGPT being in its early growth stages, this platform could escalate cybersecurity threats if not adequately safeguarded. Hence, AI could potentially have detrimental impacts on cyber-security.

Robinson (2023) brought up significant inquiries concerning ChatGPT, touching on matters related to morality, ethics, and privacy. For example, the author pondered the implications of robots becoming more human-like in appearance. If they are not acknowledged as sentient beings with rights, would we continue to identify them as machines? Robinson concluded that these kinds of questions need thoughtful deliberation before any revolutionary technology like this can be widely adopted.

Due to the intrinsic prejudices present in the training data, ChatGPT might generate biased (Biswas, 2023a) or detrimental responses. Kalla and Smith (2023) confirmed that the possibility of bias in ChatGPT's replies is a significant disadvantage. ChatGPT is educated using an extensive volume of text data, and its replies might include biases and errors. For example, ChatGPT might generate replies that are insulting or biased if the training data leans towards a specific demographic or cultural viewpoint. Addington (2023) provides instances of situations such as gender, racial, and political biases. The authors emphasized the importance of ensuring that the datasets used to educate ChatGPT are diverse and accurately depict various perspectives and groups. This is to guarantee that the replies generated by

ChatGPT are unbiased and just. Furthermore, it might be essential to develop tools for detecting and mitigating bias. Conversely, the authors also discussed the dangers of phishing attacks and data leakage associated with using ChatGPT. In phishing attacks, attackers trick users into revealing confidential information, such as usernames and passwords, using ChatGPT's conversational interface. However, unauthorized access to ChatGPT could result in data breaches and information leakage. The authors recognize that OpenAI has implemented measures, such as data encryption, access control, and security monitoring. However, the threat of cyberattacks is ever-present, and it's crucial to realize that no security mechanism is foolproof. It's vital for businesses to stay vigilant and update their security measures in response to new threats. Additionally, ChatGPT users should be aware of these potential threats and take necessary precautions to minimize risk.

Ognibene et al. (2023) highlighted that chatbots can pose risks to various facets of human existence, including identity, value, safety, uniqueness, inequality, resources, and employment. The authors proposed that when developing and implementing advanced AI systems, like ChatGPT, it's important to take into account its emotional and societal repercussions. Moreover, the implementation of suitable standards to mitigate or avoid adverse impacts is necessary. As artificial intelligence (AI) technology evolves, it becomes imperative to address pertinent societal apprehensions and manage its usage. To achieve this, a collaborative effort is needed from policymakers, specialists, experts, decision-makers, and the general public.

Yang and his team (2023) offered an exhaustive analysis of the security threats, vulnerabilities, and countermeasures associated with chatbots, and noted that there are still specific areas where further research is needed to address certain security issues. These areas encompass authorization and authentication procedures for chatbots, the identification and thwarting of harmful chatbots, ethical dilemmas, security risks tied to information security in chatbot deployment, and the influence of chatbots on social engineering attacks. Furthermore, the authors underscored some recommended practices for building secure chatbots, which include carrying out comprehensive security evaluations, implementing user authentication and authorization, employing encryption for data safeguarding, routinely updating and patching chatbots, and enlightening users about security best practices. Nonetheless, the issues brought up necessitate substantial cooperation and time.

Moreover, any security solution should be lightweight to preserve the Quality of Service (QoS) provided by the Intelligent Automation (IA) systems.

Unscrupulous individuals might leverage ChatGPT's advanced language generation abilities to gain insights about their potential victims. This could be beneficial during the initial phases of a cyberattack, where the perpetrator is collecting information about the target to strategize the most effective attack. The accumulated data can be utilized for phishing, social engineering, or to exploit known weak spots. Details about the target company's technologies, systems, organizational hierarchy, employees, challenges they encounter, and more can be compiled. A potential objective could be to create a comprehensive profile for a specific employee that includes details about their professional and personal lives, social media presence, hobbies, family, and relationships. ChatGPT can augment the process of collecting this information by offering suggestions, applying relevant statistics, and accelerating the overall process. The gathered data can be exploited in malicious activities such as identity theft, harassment, or blackmail (Derner & Batistič, 2023).

From the dawn of the Internet, cybercriminals have utilized emails as primary conduits for spam, harmful URLs, attack vectors, and other detrimental content. Cambiaso and Caviglione (2023) suggested that nearly 90% of all email traffic aids in fraudulent and illicit activities, a trend expected to persist. Consequently, mitigating the impact of damaging and unwanted emails is crucial, not just from a human perspective but also for conserving resources like storage capacity and server bandwidth. In this scenario, the researchers (Cambiaso & Caviglione, 2023) explored the potential of ChatGPT to generate email content that would lure scammers and deplete their resources. They noted that AI could serve as a practical and beneficial instrument. However, the deployment of ChatGPT has brought up numerous challenges. Factors such as human and ethical considerations, computational enhancements, and limitations in explainability must be taken into account when incorporating AI into high-quality security systems. The authors underscored several issues that emerged from their experiment, including the following:

- Distinct frameworks and conceptualizations are necessary for each category of problem. A substantial collection of messages is needed to educate the model, along with expertise in feature extraction, and significant storage and computational capacities.
- An ample amount of authentic messages need to be gathered for the AI-powered mail system to generate valuable replies. Nonetheless, this could potentially clash with legal guidelines and privacy policies.

- A fraudster might detect the absence of a human interlocutor even if the text generated by ChatGPT seems coherent and authentic, due to linguistic tendencies, an absence of syntax mistakes, or overly rapid replies.

The previously mentioned study, outlined above, offers a concise overview of the potential hazards and dangers that ChatGPT may present. It delivers cutting-edge research data that has been recently shared. Importantly, ChatGPT is a promising tool with a substantial positive impact across various domains. Nevertheless, it can be misused in a negative manner, either directly or indirectly, leading to several significant security risks that need to be addressed by implementing suitable strategies and regulations to reduce risk. For fairness, a set of questions were posed to ChatGPT to gather its responses. Table 2 enumerates some of these questions along with the responses from ChatGPT.

Table 2: Threats Exploiting ChatGPT

Questions	ChatGPT Response
Is ChatGPT posed security risks?	ChatGPT is a chatbot engineered to offer conversational aid and support to users. Although the chatbot itself doesn't present security threats, it's crucial to exercise prudence when divulging personal or sensitive data during interactions. Chatbots possess restricted abilities and are generally incapable of accessing or storing sensitive details like passwords or credit card digits. Nonetheless, it's vital to remain vigilant when utilizing any digital communication tool, as there's always a possibility that an external party could seize or access your data. To safeguard your privacy and security when using ChatGPT or any other digital communication tool, it's advised to refrain from disclosing personal, sensitive, or confidential data during interactions. Furthermore, ensure to employ robust passwords and update them frequently, and abstain from downloading or clicking on links from unverified sources.
What is the primary cybersecurity threat that ChatGPT is susceptible to?	As an AI chatbot, ChatGPT might be exposed to several cybersecurity threats, with the most prevalent being: <ul style="list-style-type: none"> • Phishing attacks - where an attacker attempts to acquire confidential data like login details, credit card numbers, or other personal data by masquerading as a trustworthy entity or service. • Malware attacks - where an attacker installs harmful software on a user's device to access data or remotely control the device. • DDoS (Distributed Denial of Service) attacks - where an attacker floods the server hosting the chatbot with an excessive amount of traffic, rendering it inaccessible. • Social engineering attacks - where an attacker deceives or manipulates individuals into revealing sensitive information. To mitigate these threats, it's crucial for ChatGPT's developers to ensure the chatbot is securely designed. Additionally, users of the chatbot must adopt safe online practices such as not divulging sensitive information to untrustworthy chatbots and avoiding suspicious or harmful links. Implementing strong encryption techniques like HTTPS and SSL/TLS is also necessary to protect any sensitive data transmitted via the chatbot.

These replies continue to highlight the potential for cyberattacks that could leverage ChatGPT's functionalities. Furthermore, these replies corroborate our observations that ChatGPT could be misused in ways that present several

significant security concerns. These issues necessitate the implementation of appropriate safeguards and protocols to minimize risk.

5. DISCUSSION AND RECOMMENDATION

ChatGPT possesses an impressive ability to produce lifelike responses across a wide array of subjects. This is a notable progression for mankind. Nonetheless, it has been predominantly portrayed in the realm of cybersecurity as a means to amplify security hazards. ChatGPT enables novices to become skilled. For instance, with just a few interactions, it's possible to acquire all the necessary elements for phishing or other forms of assault. This poses substantial cybersecurity threats that must be tackled. ChatGPT carries the danger of providing cybercriminals with straightforward access to scripting and coding, thereby effectively lowering the barriers to entry in this domain.

Despite ChatGPT's proficiency in generating programming scripts, it is designed to avoid crafting any code that appears harmful or malevolent. Nevertheless, it is unquestionably conceivable for cybercriminals to deceive ChatGPT into producing detrimental code.

ChatGPT operates as a data-centric model, and its effectiveness is directly proportional to the quality of the data it's trained on. The precision and success of any AI model are contingent upon the judicious selection of training data. In the context of ChatGPT, it's trained on a diverse array of data sources, encompassing articles, books, online discussion forums, and various other websites. Moreover, as users engage with this system, ChatGPT continually evolves and refines its responses. However, training ChatGPT on unregulated Internet data without explicit guidelines could raise concerns about the AI's ability to generate constructive responses or make suitable decisions. Consequently, the outcomes of such an uncontrolled training methodology could potentially be risky.

OpenAI, the architect of ChatGPT, has been relentless in its efforts to safeguard ChatGPT from executing harmful code by implementing a variety of measures and constraints. However, it has been observed that ChatGPT has certain limitations in adhering to these constraints, thereby providing loopholes for malicious entities to exploit. The task of ensuring the adequacy and effectiveness of these constraints remains a daunting challenge, necessitating further exploration and diligence from developers and the research fraternity.

Given that the chatbot is incapable of learning from encrypted data, the use of unencrypted data for training amplifies the risk of information exposure to unwanted parties. One of the paramount challenges in the realm of information security is the protection of sensitive user data.

There are numerous facets that demand meticulous scrutiny as they present tangible challenges to ChatGPT. These facets encompass authorization and authentication protocols for chatbots, the identification and mitigation of malevolent chatbots, ethical considerations, security hazards associated with information security in chatbot deployment, and the influence of chatbots on social engineering attacks.

Every technology, including chatbots like ChatGPT, is vulnerable to misuse by nefarious entities for their own malevolent purposes. There have been cases where cybercriminals have leveraged chatbots to initiate attacks, such as phishing schemes or spreading malware or viruses. These attacks can be orchestrated through various avenues, including vulnerabilities in the chatbot software, exploiting the potential of natural language processing algorithms, and taking advantage of weaknesses in machine learning. Perpetrators might exploit ChatGPT to fabricate platforms and applications that mimic others and provide free access to lure users. Furthermore, they might manipulate chatbots to create applications intended to collect confidential information or disseminate malware on user devices. Malevolent entities may employ ChatGPT to gather more information about their targets. ChatGPT can assist in this data acquisition process by making recommendations, utilizing pertinent statistics, and expediting the entire procedure. This information can be used for a range of illegal activities, including phishing, social engineering, and identity theft. Due to the generation of biased and misleading results by ChatGPT, which raises ethical concerns, certain countries have banned the use of ChatGPT.

In general, there are numerous obstacles in cybersecurity that stem from the AI breakthrough, ChatGPT. Given that ChatGPT is in its nascent phase, it's feasible for security experts and developers to enhance it and alleviate any potential threats. Invariably, security protocols trail behind innovation, hence the need to focus on them. Consequently, the author advises that it's imperative for chatbot developers and users to implement the necessary security measures, such as conducting comprehensive security evaluations, frequently updating the software, implementing strong authentication procedures, employing encryption for data safeguarding, enlightening users about security best practices, and monitoring unusual behavior, to avert such attacks.

Conversely, ChatGPT could serve as an instrument for researchers to better understand and predict various types of attacks, ultimately resulting in more efficient cybersecurity defenses. Therefore, it's crucial to leverage these features to establish policies to prevent the abuse and exploitation of generative AI.

6. CONCLUSION

Conventional security protocols are inadequate to counter attacks due to the rise of advanced cyber threats. AI provides a way to enhance cybersecurity defenses by automating the identification and mitigation of threats, scrutinizing enormous amounts of data, and predicting potential attacks.

Nonetheless, AI systems present a conundrum for cybersecurity. ChatGPT excels at generating believable responses across a broad spectrum of subjects. It's an impressive innovation that has recently garnered interest; however, it is predominantly portrayed in the context of cybersecurity as a tool that boosts cyberattack capabilities.

This article discusses numerous harmful activities that ChatGPT can facilitate or amplify, including phishing, social engineering, malware, privacy violations, enabling novice hackers, hacking, and creating malicious code.

Hence, developers of ChatGPT must ensure that the chatbot is designed securely to counter such attacks, and users of chatbots must practice safe online behaviors, such as not sharing personal or sensitive information with chatbots and avoiding unfamiliar or risky links. Robust encryption methods should also be implemented to safeguard the sensitive information transmitted via chatbots.

In conclusion, it is imperative for chatbot creators and users to implement necessary security measures to ward off such attacks. This includes conducting comprehensive security evaluations, frequently updating the software, implementing strong authentication procedures, utilizing encryption for safeguarding data, enlightening users about safe online practices, and monitoring for unusual activities. Furthermore, it is advisable to foster a productive collaboration to devise security protocols and rules that deter such attacks without imposing additional burden, thereby preserving the QoS of this promising tool.

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCES

- Addington, S. (2023). ChatGPT : Cyber Security Threats and Countermeasures.
- Ali, H., & Aysan, A. F. (2023). What will ChatGPT Revolutionize in Financial Industry? [3] Aljanabi, M. (2023). ChatGPT: Future directions and open possibilities. *Mesopotamian journal of cybersecurity*, 2023, 16-17.
- Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R.J., Esmaeili, M., Majdabadkohne, R.M., & Pasehvar, M (2023). ChatGPT : Applications, Opportunities, and Threats. In *IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 274-279.
- Biswas, S. (2023). Prospective Role of Chat GPT in the Military: According to ChatGPT. *Qeios*. [6] Biswas, S. (2023b). Role of Chat GPT in Education. *J of ENT Surgery Research*, 1(1), 01-03.
- Biswas, S. (2023c). Role of ChatGPT in Cybersecurity.
- Cambiaso, E., & Caviglione, L. (2023). Scamming the Scammers: Using ChatGPT to Reply Mails for Wasting Time and Resources.
- Charan, P.V., Chunduri, H., Anand, P.M., & Shukla, S.K. (2023). From Text to MITRE Techniques: Exploring the malicious Use of Large Language Models for Generating Cyber Attack Payloads.
- Choi, J. H., Hickman, K. E., Monahan, A., & Schwarcz, D. (2023). ChatGPT goes to law school *Journal of Legal Education* (Forthcoming).
- Derner, E., & Batistič, K. (2023). Beyond the Safeguards: Exploring the Security Risks of ChatGPT.
- Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., & Wright, R. (2023). "So, What if ChatGPT wrote it?" Multidisciplinary perspective on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information management*. 71.

- Eggmann, F., Weiger, R., Zitzmann, N.U., & Blatz, M.B. (2023). Implications of large language models such as ChatGPT for dental medicine. *Journal of Esthetic and Restorative Dentistry*, 1098-1102.
- Elsadig, M. A. (2023). Detection of Denial-of-Service Attack in Wireless Sensor Networks: A Lightweight Machine Learning Approach. *IEEE Access*, 11, 83537-83522.
- Elsadig, M.A., & Gafar, A. (2022). Covert channel detection: machine learning approaches. *IEEE Access*, 10, 38391-38405.
- Esmailzadeh, Y. (2023). Potential Risks of ChatGPT: Implications for Counterterrorism and International Security. *International Journal of Multicultural and Multireligious Understanding (IJMMU)*, 10.
- Gabriela, T.R., & Axinte, S.D. (2023). ChatGPT-Information Security Overview. In *International Conference on Cybersecurity and Cybercrime*, 10, 81-85.
- George, A.S., & George, A.H. (2023). A review of ChatGPT AI's impact on several business sectors. *Partners Universal International Innovation Journal*, 1(1), 9-23.
- Gill, S.S., & Kaur, R. (2023). ChatGPT: Vision and challenges. *Internet of Things and Cyber- Physical Systems*, 3, 262-271.
- Grbic, D.V., & Dujlovic, I. (2023). Social engineering with ChatGPT. In *IEEE 22nd International Symposium Infoteh-Jahorina (Infoteh)*, 1-5.
- Kalla, D., & Smith, N. (2023). Study and Analysis of Chat GPT and its Impact on Different Fields of Study. *International Journal of Innovative Science and Research Technology*, 8(3), 827-833.
- Khosravi, H., Shafie, M.R., Hajiabadi, M., Raihan, A.S., & Ahmed, I. (2023). Chatbots and ChatGPT: A bibliometric analysis and systematic review of publications in Web of Science and Scopus databases.
- Khoury, R., Avila, A.R., Brunelle, J., & Camara, B.M. (2023). How Secure is Code Generated by ChatGPT?
- Kshetri, N. (2023). ChatGPT in developing economies. *IT Professional*, 25(2), 16-19.
- Liu, B., Xiao, B., Jiang, X., Cen, S., He, X., & Dou, W. (2023). Adversarial Attacks on Large Language Model-Based System and Mitigating Strategies: A Case Study on ChatGPT. *Security and Communication Networks*, 2023, 1-10.
- Liu, C., Bao, X., Zhang, H., Zhang, N., Hu, H., Zhang, X., & Yan, M. (2023). Improving ChatGPT Prompt for Code Generation.
- Mansfield-Devine, S. (2023). Weaponising ChatGPT. *Network Security*, 2023(4).
- Marshall, J. (2023). What Effects Do Large Language Models Have on Cybersecurity.
- McKee, F., & Noever, D. (2023). Chatbots in a Honeypot World.
- Nair, M., Sadhukhan, R., & Mukhopadhyay, D. (2023). Generating secure hardware using chatgpt resistant to cwes. *Cryptology ePrint Archive*, 1-16.
- Ognibene, D., Baldissarri, C., & Manfredi, A. (2023). Does ChatGPT pose a threat to human identity?
- O'Rourke, M. (2023). Chatgpt poses cybersecurity threats. *Risk Management*, 70(2), 3030.
- Rahman, M.M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 1-21.
- Ray, P.P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and CyberPhysical Systems*, 3, 121-154.
- Robinson, J. (2023). The cost of science a look at the ethical implications of chatgpt.
- Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet*, 15(6), 1-24.
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare*, 11(6), 1-20.
- Sebastian, G. (2023). Do ChatGPT and other AI chatbots pose a cybersecurity risk? An exploratory study. *International Journal of Security and Privacy in Pervasive Computing (IJSPPC)*, 15(1), 1-11.
- Sebastian, G. (2023). Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information.
- Sharma, P., & Dash, B. (2023). Impact of big data analytics and ChatGPT on cybersecurity. In *IEEE 4th International Conference on Computing and Communication Systems (I3CS)*, 1-6.
- Shoufan, A. (2023). Exploring Students' Perceptions of CHATGPT: Thematic Analysis and Follow-Up Survey. *IEEE Access*, 11, 38805-38818.
- Sun, W., & Yao, J. (2023). Exploring the Potential Application of ChatGPT in Preparing for ABET Accreditation.
- Surameery, N.M.S., & Shakor, M.Y. (2023). Use chat gpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering (IJITC)*, 3(01), 17-22.
- Temara, S. (2023). Maximizing penetration testing success with effective reconnaissance techniques using chatgpt.

- Thorncharoensri, P., Susilo, W., & Baek, J. (2019). Efficient Controlled Signature for a Large Network with Multi Security-level Setting. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 10(3), 1-20.
- Xames, M.D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. *Journal of Applied Learning and Teaching*, 6(1), 390-395.
- Yang, J., Chen, Y. L., Por, L. Y., & Ku, C. S. (2023). A systematic literature review of information security in chatbots. *Applied Sciences*, 13(11), 1-18.