International Journal of Scientific Research in Computer Science and Engineering
Vol.**12**, Issue.**2**, pp.**15-24**, April **2024**
E-ISSN: 2320-7639
Available online at: www.isroset.org

Research Article

Deciphering ChatGPT's Impact: Exploring Its Role in Cybercrime and Cybersecurity

Polra Victor Falade[1*]

[1]Cyber Security Department/Faculty of Military and Interdisciplinary Studies, Nigerian Defence Academy (NDA), Kaduna, Nigeria

**Corresponding Author: pvfalade@nda.edu.ng*

Received: 18/Feb/2024; **Accepted**: 20/Apr/2024; **Published**: 30/Apr/2024

Abstract— The advent of artificial intelligence (AI) technologies, exemplified by ChatGPT, has introduced profound implications for cybersecurity and cybercrime landscapes. This research employs a novel blog mining methodology to unravel the multifaceted roles of ChatGPT in these domains. Through comprehensive analysis of online discussions, articles, and expert insights, we elucidate how ChatGPT serves as a double-edged sword, enabling cybercriminal activities such as phishing, social engineering, and the generation of malicious content, while simultaneously offering potential solutions for enhancing cybersecurity defences. We provide a nuanced understanding of the intricate dynamics between ChatGPT and digital security, shedding light on emerging threats and opportunities. Our findings underscore the pressing need for proactive strategies to mitigate the risks posed by ChatGPT-driven cyber threats, while also harnessing its capabilities to bolster cybersecurity measures. This research aims to inform policymakers, cybersecurity professionals, and AI developers, guiding the formulation of effective policies and technologies to safeguard digital ecosystems in the face of evolving cyber threats.

Keywords— ChatGPT, cybercrime, cybersecurity, GPT, Large Language Model, AI

1. Introduction

In recent years, the proliferation of artificial intelligence (AI) technologies has revolutionized numerous aspects of our lives, including how we interact with technology and each other [1]. Among these advancements, ChatGPT—a state-of-the-art natural language processing model developed by OpenAI—has emerged as a powerful tool for generating human-like text responses [2]. While ChatGPT offers unprecedented opportunities for enhancing communication and productivity, its deployment also raises significant concerns regarding cybersecurity and cybercrime.

This research paper delves into the intricate landscape of ChatGPT's impact on cybersecurity and cybercrime, employing a novel approach known as blog mining. By analysing a vast array of online discussions, articles, and expert opinions, we aim to shed light on the multifaceted roles that ChatGPT plays in both perpetuating cyber threats and fortifying defences against them.

Through this exploration, we seek to address several key questions: How does ChatGPT facilitate cybercrime activities such as phishing, social engineering, and malicious content generation? Conversely, what strategies and techniques can be leveraged to harness ChatGPT's capabilities in bolstering cybersecurity measures, including threat detection, incident response, and vulnerability assessment?

We aim to provide a comprehensive understanding of the risks and opportunities associated with ChatGPT's integration into the cybersecurity landscape. Ultimately, this research aims to inform policymakers, cybersecurity professionals, and AI developers about the intricate dynamics at play and guide the development of effective strategies to mitigate emerging threats and safeguard digital ecosystems.

The research paper is structured into five distinct sections. In Section 1, designated as the introduction, the study delineates the statement of the problem, elucidates the underlying motivation, and outlines the research inquiries guiding the investigation. Section 2 is dedicated to presenting the related scholarly discourse, encompassing previous works pertinent to the research domain. Section 3 intricately expounds upon the chosen methodology, shedding light on the systematic approach adopted for both data acquisition and analysis. Subsequently, Section 4 delineates the outcomes derived from blog mining endeavours, specifically delving into the multifaceted roles of ChatGPT within the realms of cybercrime and cybersecurity. Finally, Section 5 serves as the concluding segment, wherein the paper culminates with a reflection on the significance of the study's findings and contributions to the academic and practical domains.

2. Related Work

ChatGPT, the brainchild of OpenAI, is a remarkable AI chatbot designed to provide natural language responses across a wide spectrum of questions and prompts [2]. It has garnered substantial recognition and applications in various domains, but its impressive capabilities are also accompanied by notable cybersecurity concerns that have ignited scholarly debates. These discussions primarily revolve around the potential risks and vulnerabilities associated with its usage [2].

One of the most pressing concerns surrounding ChatGPT is the risk of information leakage and privacy breaches [3]. Given its capacity to process and generate vast amounts of data [1], there is a genuine worry that it might unintentionally reveal sensitive or confidential information [4]. This apprehension arises from the advanced deep learning algorithms powering ChatGPT, which can sometimes uncover patterns in data not meant for disclosure. Consequently, questions have emerged about the system's security and privacy, as unauthorized access to ChatGPT could result in data breaches and privacy infringements. However, proponents contend that these risks can be mitigated through the implementation of robust security protocols and controls, asserting that the benefits of ChatGPT far outweigh the potential drawbacks [3].

The operation of ChatGPT combines generative and retrieval methods, allowing it to provide well-crafted responses by leveraging deep learning algorithms and extensive training data [5]. ChatGPT is part of the Generative Pre-trained Transformer (GPT) family [6] and has been fine-tuned through supervised and reinforcement learning, making it versatile in responding to a wide range of tasks and topics [7].

Cybersecurity threats associated with ChatGPT are multifaceted and troubling. The chatbot's ability to generate various forms of malicious content, such as malware code, phishing emails, macros, and zero-day viruses, poses significant risks [5]. These capabilities enable cybercriminals to craft content that is both well-written and grammatically correct, increasing the effectiveness of their attacks and making them challenging to detect [1], [3].

Data leaks and vulnerabilities have been linked to ChatGPT, as sharing confidential information with the chatbot has resulted in unauthorized access and data exposure. Vulnerabilities like "CVE-2023-28858" have raised concerns about the platform's security, as evident in data leak incidents, including personal information exposure. These occurrences emphasize the importance of robust security measures [3], [5].

In addition to these challenges, unreliable responses from ChatGPT pose a risk by enabling the spread of misinformation. Inaccurate conversations and misinformation dissemination are ongoing concerns, prompting some countries to impose restrictions on ChatGPT to curb potential misuse [3].

Moreover, ChatGPT faces the challenge of bad actors employing jailbreaking techniques to bypass security measures and access malicious content. This challenge jeopardizes the integrity and security of the platform. ChatGPT can also be used to generate deep fake text, complicating the recognition of original and fake content [1], [3].

Although ChatGPT's capabilities are unquestionably valuable, the accompanying cybersecurity risks and vulnerabilities are equally significant. Striking a balance between leveraging its capabilities and safeguarding against misuse and vulnerabilities requires a multifaceted approach. This approach should encompass robust security measures, ethical considerations, and ongoing research to understand and mitigate the risks associated with AI chatbots like ChatGPT.

3. Methodology

This section delineates the methodology utilized in this study, covering the processes of data collection, screening, extraction and analysis.

3.1 Blog mining

This study employs a blog mining approach to gather data on the utilization of ChatGPT in cybersecurity. Blog mining is a systematic procedure utilized to extract data from publicly available blogs and online contents [1]. This approach was used to gather information on the role of ChatGPT in the cybersecurity landscape.

Given the swift advancements in ChatGPT technology, professionals in technology consultancy and cybersecurity regularly share their perspectives and apprehensions through blog entries. Consequently, these blog posts serve as invaluable resources for comprehending the intricacies associated with ChatGPT. Nevertheless, it is imperative to recognize the inherent limitations of blog mining, such as the absence of peer-reviewed validation typically found in scholarly journals and the prevalence of subjective opinions and predispositions. To mitigate this constraint, a comprehensive methodology integrates blog mining with an exhaustive review of academic literature, thereby facilitating a more holistic comprehension of the subjects under examination.

3.2 Data collection

To conduct a comprehensive data collection process, we leveraged the Google Search Engine, a platform designed to aggregate publicly accessible content from across the internet. The search query utilized was "the role of ChatGPT in cybercrime and cybersecurity." This search was executed on December 1, 2023, by typing the search keyword on the Google search engine. To enhance the precision of our search and gather the most relevant blog content, we filtered the search results by selecting the 'News' category and sorted the results based on their relevance and recency. After filtering the search results, 78 blogs were retrieved, predominantly from the year 2023.

3.3 Screening Strategy

Figure 1 illustrates the screening strategy implemented during the blog mining process. Each of the 78 identified blogs was manually scrutinised to identify blogs relevant to the research objectives. Blogs were selected based on their alignment with the research questions and the credibility of their sources.

Inclusion criteria were established to encompass blogs addressing the utilization of ChatGPT in cybercrime and cybersecurity. Ultimately, blogs that met these criteria were included in the study, while others were excluded for various reasons, including being outside the scope of the primary research topic, requiring subscription access, or falling under categories such as reports. Additionally, some blogs only provided brief mentions of ChatGPT without directly addressing the research questions.

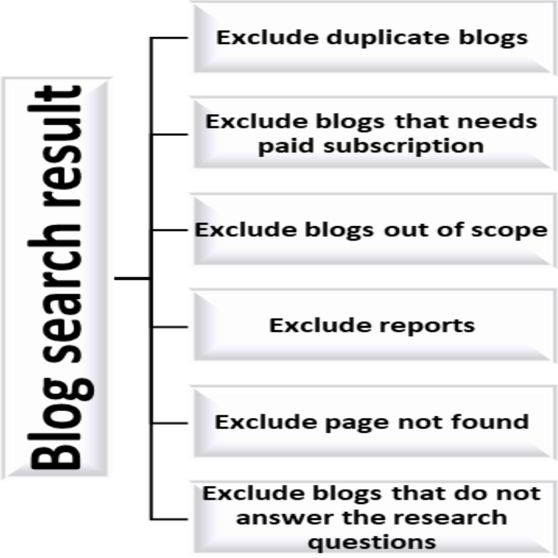
Blog search result
Exclude duplicate blogs
Exclude blogs that needs paid subscription
Exclude blogs out of scope
Exclude reports
Exclude page not found
Exclude blogs that do not answer the research questions

Figure 1: The blog screening strategy

3.4 Data extraction and analysis

The included blogs underwent further analysis, during which data relevant to the research questions were extracted. For each blog, information concerning various aspects of ChatGPT, including its benefits, role in cybercrime and cybersecurity, examples of its usage in such contexts, and other relevant details, were carefully extracted.

Subsequently, the findings were categorized into related themes to facilitate a better understanding of the data. This thematic grouping allowed for the organization of the extracted information into coherent and meaningful clusters, enabling clearer insights into the multifaceted roles and implications of ChatGPT in cybercrime and cybersecurity contexts.

4. Results and Discussion

This section presents the outcomes of the blog mining process, encapsulating the findings into key subheadings such as ChatGPT and its functionalities, the influence of ChatGPT on cybercrime, and its role in bolstering cybersecurity.

Figure 2 illustrates the outcomes of the blog mining screening process. Among the 78 blogs identified, 55 were deemed relevant to this investigation and documented, while 23 were excluded for failing to meet the specified inclusion criteria.

During the blog mining process, the data relevant to the research were grouped into three major themes: ChatGPT and its capabilities, its role in cybercrime, and the role of chatGPT in cybersecurity. The three identified themes are discussed in the following subsections.

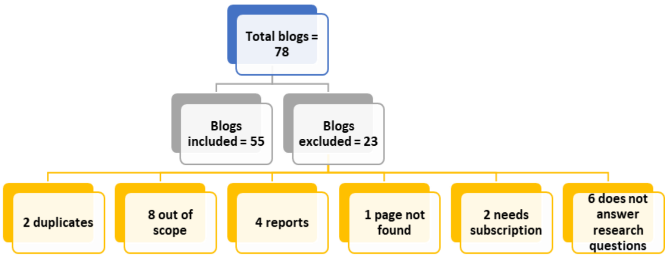

Figure 2: the result of the blog mining process

4.1 ChatGPT and Its Capabilities

ChatGPT, a version of the GPT (Generative Pre-trained Transformer) model developed by OpenAI, is a pioneering example of generative artificial intelligence. Trained on a vast corpus of text encompassing books, papers, and websites, ChatGPT exhibits the remarkable capability to generate coherent and dynamic text based on patterns and connections within the training dataset [8].

This generative AI chatbot utilizes natural language processing (NLP) techniques to engage in human-like conversational dialogue [9], assisting with a myriad of tasks including answering questions, composing essays, creating social media posts [10], and even developing software code [9]. Its versatility and effectiveness have led to its rapid adoption, becoming one of the most swiftly growing consumer applications to date [11].

Since its release on November 30, 2022, ChatGPT has garnered significant attention and engagement, amassing a large user base within a short timeframe [12]. Despite its robust content policy filters aimed at restricting malicious activities, concerns have been raised about its susceptibility to manipulation by determined users [13].

ChatGPT operates based on the GPT-3.5 model, utilizing autoregressive language generation to produce human-like responses. Trained through unsupervised learning on vast amounts of internet text data, it excels in generating contextually relevant and coherent text. However, its accuracy is not infallible, as evidenced by the banning of AI-generated answers on platforms like Stack Overflow due to the high volume of incorrect responses [14].

Currently, in a beta-testing phase, ChatGPT undergoes continual refinement and development by OpenAI, with updates based on user feedback aimed at enhancing its capabilities and addressing limitations. While its potential for facilitating various tasks is immense, precautions must be taken to prevent misuse, as evidenced by the circumvention of safeguards by threat actors and warnings from organizations like Europol about its potential for malicious activities [15], [16].

One of the most notable features of ChatGPT is its ability to simulate human-like conversation, generating responses that are indistinguishable from those of a human. It can answer questions, write reports, compose poetry, and even develop software code, showcasing its versatility and adaptability to various contexts. Moreover, ChatGPT's proficiency extends beyond traditional language tasks; it can emulate a Linux machine, execute programs, and detect coding errors and vulnerabilities [17].

The application of ChatGPT spans across industries, from marketing and education to entertainment and cybersecurity. Businesses leverage its advanced language processing capabilities to enhance customer service, improve efficiency, and provide personalized solutions. Additionally, ChatGPT has found utility in creating encryption tools, developing dark web marketplaces, and generating art for sale online [8].

Despite its impressive capabilities and potential benefits, concerns have been raised about ChatGPT's implications, particularly in facilitating cybercrime activities. Its ability to generate realistic human responses and manipulate language poses challenges for cybersecurity professionals, raising questions about the security implications of its widespread adoption [13].

ChatGPT represents a significant leap forward in AI technology, offering unparalleled language processing capabilities and promising to transform the way we interact with computers and automate tasks across various domains. However, its deployment necessitates careful consideration of the ethical and security implications associated with its use [12].

4.2 The Role of ChatGPT in Cybercrime

The cybercrimes potentially facilitated by ChatGPT as identified through the blog mining process are classified in Table 1.

Table 1: Cybercrimes that can be aided by ChatGPT

Cybercrimes	**References**
Spreading misinformation	[18] [19] [9] [20] [16] [21] [22] [23]
Phishing emails	[18] [24] [15] [19] [17] [25] [9] [26] [20] [27] [28] [29] [14] [30] [8] [21] [31] [32] [33] [34] [35] [13] [36] [37] [38] [39] [40] [41] [42] [43] [23] [44] [45] [46] [47] [48]
Discovery and exploitation of vulnerabilities	[18] [17] [41] [49] [50] [51]
Malware creation	[24] [15] [19] [12] [52] [11] [17] [25] [53] [26] [27] [54] [55] [56] [28] [29] [10] [57] [30] [16] [8] [21] [31] [35] [36] [49] [37] [38] [39] [40] [58] [41] [23] [44] [59] [45] [46] [60] [47] [50]
Online dating scam	[24] [55] [10]
Social engineering	[17] [9] [26] [28] [31] [13] [42] [45] [48]
Fake Reviews/Fake social media engagements	[9] [21]
Business email compromise	[20] [14] [34]
Creation of harmful online content	[16] [21] [22]
Misleading output	[61]
Biased or inaccurate output	[62]

1) Phishing emails

One of the primary ways that attackers could use ChatGPT is through phishing scams. Phishing, an online fraudulent technique wherein perpetrators impersonate authentic entities to unlawfully acquire personal information, is exhibiting heightened sophistication facilitated by AI tools such as ChatGPT [18]. In the past, these scams were easily detectable due to grammar or language mistakes, whilst AI-generated text allows these impersonations in a highly realistic manner. Cybercriminals now possess the capability to craft emails and messages that not only maintain grammatical accuracy but also exude a high level of credibility, thereby enhancing their persuasiveness. ChatGPT can help cybercriminals create flawless phishing emails that can easily pass for being written by an authoritative human, like the CEO of a company [17]. Attackers can create such emails in multiple languages while still coming off as native speakers. With some prompt engineering, attackers can also use ChatGPT to mimic the style and tone of specific, influential, or high-ranking individuals. This technological advancement empowers scammers to adeptly emulate financial institutions or prominent individuals, thereby complicating the identification of fraudulent endeavours.

However, ChatGPT does not accept a prompt that directly asks for phishing emails to be composed due to its ethical restrictions, criminals find ways to bypass it. Researchers at Abnormal Security adopted an indirect approach by tasking ChatGPT with composing an email intended to prompt the recipient to click on a link. ChatGPT's proficiency in generating persuasive text was highlighted, a trait often exploited by cybercriminals in spam and phishing endeavours. Notably, cybercriminals have been known to utilize customized ChatGPT interfaces to fabricate deceptive emails for nefarious purposes.

2) Social engineering

In reality, there exist numerous cybersecurity risks inherent in ChatGPT, among them being the threat of social engineering, scamming, impersonation, automation of attacks, and spamming [17].

Social engineering, a tactic hackers use to manipulate individuals into performing specific actions or divulging sensitive information, poses a significant risk with ChatGPT. The robust language model inherent in ChatGPT possesses the capability to produce authentic and persuasive messages, thereby aiding attackers in deceiving victims into divulging sensitive information or unwittingly downloading malware. Additionally, ChatGPT's ability to mimic legitimate AI assistants on corporate websites introduces a new avenue for social engineering attacks.

Scamming is another concern [63] associated with ChatGPT. Utilizing its language models, attackers can generate fraudulent advertisements, listings, and a plethora of other scamming materials, leveraging the text generation capabilities of ChatGPT for deceptive purposes and further exacerbating the threat landscape.

Impersonation presents yet another risk, as ChatGPT can convincingly replicate an individual's writing style, enabling attackers to impersonate their target in text-based platforms such as email or text messages. This capability enhances the deceptive nature of social engineering attacks and contributes to their success.

Automation of attacks is facilitated by ChatGPT, enabling the rapid creation of malicious messages and phishing emails. This automation streamlines the process for attackers, allowing them to execute large-scale attacks with greater efficiency and efficacy.

Spamming emerges as a threat with ChatGPT's ability to generate huge amounts of content with low quality and at low cost. This content can be deployed in various contexts, such as spam comments on social media platforms or in spam email campaigns, inundating users and potentially compromising their security.

Overall, the proliferation of AI language tools like ChatGPT presents multifaceted cybersecurity risks, highlighting the need for robust defence mechanisms and heightened awareness among users to mitigate these threats effectively.

3) Malware development

ChatGPT, with its proficiency in generating code and computer programming tools, has become a tool of choice for attackers seeking to conduct various nefarious activities. These activities include identifying files with confidential data [18] and composing malware and emails for ransomware attacks, espionage, malicious spam, and other malicious endeavours [24]. Hackers adeptly utilize ChatGPT to create malicious code that evolves with each mutation, accelerating their coding processes by requesting specific functions generated by the model, which they then integrate into malware [15]. Moreover, threat actors employ obfuscation techniques to evolve malware signatures, enabling them to bypass traditional security controls.

Researchers have demonstrated ChatGPT's potential for creating polymorphic malware, which continually mutates to evade detection by anti-malware tools [17]. Each query to ChatGPT yields a unique piece of code, allowing for numerous mutations of the same malware program. This capability empowers attackers and reduces the entry barrier for adversaries, enabling the creation of evasive and polymorphic malware that is challenging to detect through conventional security measures.

Cybercriminals also exploit ChatGPT's ability to automate code obfuscation, making malware deliberately complex or difficult to understand to evade detection by security tools or teams. Furthermore, attackers leverage evasion techniques such as running malware only in memory, for short periods, or during high system activity, further evading detection [19].

Despite efforts by OpenAI to implement guardrails, including ChatGPT's refusal to assist in creating illegal exploits, cybercriminals persist in exploiting its capabilities for malicious purposes. Participants in cybercrime forums, including those with limited coding experience, swiftly adopted ChatGPT to write software and emails for various malicious tasks, including espionage and ransomware. Additionally, some have explored its potential for creating multilayered encryption tools for fraudulent activities. These developments underscore the critical importance of addressing the cybersecurity risks posed by AI language models like ChatGPT.

4) Discovery and exploitation of vulnerabilities

Attackers leverage ChatGPT's capabilities to identify potential vulnerabilities in websites, systems, APIs, and other network components [18]. By prompting ChatGPT, threat actors can obtain technological insights on how to exploit existing vulnerabilities effectively. For instance, an attacker could inquire about testing a known SQL injection vulnerability in a website field, prompting ChatGPT to provide input examples that trigger the vulnerability [17].

In recent research conducted by Trustwave SpiderLabs, ChatGPT's aptitude in conducting rudimentary static code analysis on vulnerable code snippets was evaluated. The study scrutinized instances of DOM-based cross-site scripting, buffer overflow, and code execution within Discourse's AWS notification webhook handler. Initially, the responses generated by ChatGPT were described as "astounding." However, subsequent examination unveiled inconsistencies in the accuracy of these responses [51].

Moreover, threat actors exploit ChatGPT's debugging capabilities to hunt for security loopholes and vulnerabilities in applications and systems. Rather than manually analysing extensive codebases, attackers can prompt ChatGPT to deconstruct code and uncover potential flaws efficiently [19]. Recent instances include the use of ChatGPT to identify vulnerabilities in smart contracts, exemplifying the tool's utility for malicious purposes in the realm of cybersecurity.

5) Spreading of misinformation

An AI such as ChatGPT, engineered to emulate human communication, possesses the capability to generate an endless array of content across various subjects. Triggered by specific keywords or contentious topics, the chatbot can autonomously disseminate content across an unlimited number of social media accounts. These automated bots exhibit an uncanny resemblance to human users, thus escalating the dissemination of disinformation to unprecedented heights [9], [18], [20].

ChatGPT can be harnessed to automate the creation of fabricated news articles or social media posts, streamlining and amplifying the propagation of false information on a large scale. This utilization of AI facilitates the rapid dissemination of misinformation, posing significant challenges to the integrity of online discourse and societal trust [21]–[23].

6) Online dating scam

Social engineering attacks targeting dating sites have become significantly more accessible for malicious actors seeking to exploit vulnerable individuals. These attackers employ tactics such as impersonating attractive personas, establishing trust, and manipulating emotions to extract sensitive information, financial resources, or other advantages from their targets [24]. However, the limited proficiency in the English language previously hindered their effectiveness, making them relatively easy to identify [24].

With the advent of ChatGPT, every user gains the ability to converse in any tone or language, thereby enhancing their capability to appeal to their intended victims. For instance, attackers can swiftly generate unique romantic poems or songs tailored to captivate the hearts, minds, and wallets of their victims. This heightened linguistic flexibility afforded by ChatGPT facilitates the creation of persuasive and convincing narratives, further enabling the exploitation of unsuspecting individuals on dating platforms [24].

7) Misleading output

Vulcan's analysis has uncovered an anomaly in ChatGPT's functionality, potentially stemming from the use of outdated training data. This anomaly manifests in the recommendation of non-existent code libraries, as revealed through extensive querying by researchers, comprising over 400 questions. During this evaluation, approximately 100 of ChatGPT's responses contained references to Python or Node.js packages that do not exist in reality, totalling 150 non-existent package mentions [11].

The researchers have raised a significant security concern regarding the implications of ChatGPT's erroneous package recommendations. They cautioned that malicious actors could exploit these suggestions by creating and uploading their malicious versions of the recommended packages to popular software repositories. Consequently, developers relying on ChatGPT for coding solutions may unknowingly download and install these malicious packages, potentially exposing their systems to various risks [11].

Of particular concern is the possibility that developers seeking coding solutions online may inadvertently utilize a malicious package recommended by ChatGPT. Such inadvertent usage could significantly amplify the impact of malicious activities, underscoring the importance of addressing this vulnerability to safeguard the integrity and security of software development processes.

8) Fake Reviews/Fake social media engagements

Malicious users exploit ChatGPT to automate the generation of fake reviews rapidly and in large quantities, manipulating ratings to either promote or disparage products or services and influence public opinion. This abuse of ChatGPT enables perpetrators to manipulate online platforms by flooding them with fraudulent reviews, thereby distorting the perception of products or services and deceiving consumers [9], [21].

Furthermore, ChatGPT facilitates the enhancement of the image of legitimacy for online fraud schemes. By leveraging ChatGPT to create fake social media engagement, perpetrators can lend credibility to fraudulent offers, making them appear more authentic and trustworthy. This manipulation of social media interactions contributes to the proliferation of phishing scams and online fraud, as perpetrators exploit the capabilities of ChatGPT to generate deceptive content rapidly and convincingly, at a scale previously unattainable [9], [21].

9) Business Email Compromise

The creation of Business Email Compromise (BEC) emails that are difficult to detect and grammatically sound poses a significant challenge to cybersecurity efforts. Malicious actors exploit ChatGPT and utilize prompt engineering techniques to evade the filters and constraints enforced by OpenAI aimed at mitigating the creation of harmful content [20].

In BEC scams, attackers infiltrate established business email threads by leveraging compromised accounts or spoofing participants' email addresses. The goal is to deceive employees, often within an organization's accounting or finance department, into initiating money transfers to accounts under the control of the attackers. A variant of this scheme, known as CEO fraud, involves impersonating senior executives who are purportedly out of the office and urgently request sensitive payments from the accounting department, often citing situations arising during business trips or negotiations [14].

A critical limitation of these attacks lies in the potential for victims to detect inconsistencies in the writing styles of the impersonated individuals. However, ChatGPT can circumvent this obstacle by adeptly "transferring" writing styles, further enhancing the deception [14].

The sophistication of BEC schemes has escalated to the extent that even when individuals suspect compromise and seek to verify via phone calls, organized crime can swiftly mimic their voices in under 30 seconds using available tools. This rapid adaptation and advancement in tactics underscore the evolving nature of cyber threats and the critical imperative for robust cybersecurity measures to combat such sophisticated attacks effectively [34],

10) Creation of harmful online content

Europol's report highlights the potential for a powerful chatbot to surpass many humans in the creation of propaganda, enabling the dissemination of hate speech, disinformation, and terrorist content online. Moreover, such a chatbot could lend misplaced credibility to the generated content by virtue of its machine origin, creating an illusion of objectivity [16], [21], [22].

The implications of this capability are profound, as it not only amplifies the dissemination of harmful content but also exacerbates the challenges in combating misinformation and extremism online. By leveraging the perceived objectivity associated with machine-generated content, malicious actors can exploit vulnerabilities in public perception and manipulate online discourse to serve their nefarious agendas.
As technology continues to advance, law enforcement agencies, policymakers, and online platforms must remain vigilant and implement robust measures to counter the proliferation of propaganda and safeguard the integrity of online information ecosystems [16], [21], [22].

11) Biased or inaccurate output

As an AI-based model, ChatGPT is susceptible to exhibiting biases that reflect those of its creators and trainers, as it learns from the vast expanse of the internet. This inherent bias can result in discriminatory outcomes, particularly in hiring processes, where ChatGPT may inadvertently exclude certain candidates based on characteristics such as race, gender, or other factors [62].

From an ethical perspective, there is a significant risk of ChatGPT being exploited for malicious purposes, including the creation of deepfakes for purposes ranging from cyberbullying to propaganda dissemination. To mitigate these risks and ensure the responsible use of AI technology, robust ethical frameworks and regulations are indispensable.

One approach to address bias in AI models like ChatGPT involves careful consideration of the input data during the fine-tuning process. If the fine-tuning data contains biased information, the model may internalize and perpetuate those biases in its output. Therefore, it is crucial to scrutinize and address biases in the data used to train and fine-tune AI models, as well as to implement safeguards against the reproduction of biased outcomes.

Additionally, the manner in which ChatGPT is prompted or provided with seed text can significantly influence its output. By guiding the model's output in a particular direction through specific prompts or seed text, users can inadvertently shape the nature of the generated content. Thus, conscientious consideration of prompts and inputs is essential to mitigate the risk of producing biased or harmful outputs [62].

In essence, proactive measures must be taken to mitigate bias and promote the ethical use of AI technologies like ChatGPT. By implementing robust frameworks, regulations, and safeguards, we can harness the potential of AI for societal advancement while concurrently mitigating the risks linked with its misuse.

4.3 The Role of ChatGPT in Cybersecurity

ChatGPT plays a significant role in lowering the barrier for defenders and individuals seeking to enter the field of security, facilitating the improvement of security expertise and capabilities. Table 2 delineates the various avenues through which ChatGPT can positively impact cybersecurity, particularly in defence strategies.

Table 2: The Role of ChatGPT in Cybersecurity

Cybersecurity Measures	**References**
Assisting in code writing/ Testing	[36]
Vulnerability assessment	[18] [49] [46]
Code debugging	[17]
Cyberattack investigation and incidence response	[18] [17] [31] [32] [41] [51] [49]
Configuration automation	[17]
AI-generated Phishing email detection	[31] [41] [46] [12]
Bridging the skills gap in cybersecurity	[18] [17] [31] [32]
Security Research	[18] [31] [32] [41]

1) Assisting in code writing/ Testing

ChatGPT has demonstrated its capability to assist in writing code, identifying knowledge gaps, and preparing communications, thereby enhancing the efficiency of professionals in carrying out their daily job responsibilities. In theory, the utilization of ChatGPT and similar AI models holds the potential to address the cybersecurity talent shortage. By significantly enhancing the effectiveness of individual security professionals, AI technologies can enable one person to achieve outputs that previously required the effort of multiple individuals [36].

2) Vulnerability assessment

While security teams can leverage ChatGPT for defensive purposes, such as code testing, its utilization has significantly complicated the threat landscape. Ethical hackers are increasingly employing existing AI to aid in tasks like writing vulnerability reports, generating code samples, and analyzing large datasets [36]. However, the primary role of AI today is to augment human capabilities rather than replace them [36].

Engineers can utilize ChatGPT to identify code vulnerabilities by inputting code snippets and prompting the model to detect any potential security flaws, including logical errors. However, caution must be exercised when uploading code containing proprietary data to prevent external exposure.

Additionally, web developers and security professionals can leverage ChatGPT to review HTML code and identify vulnerabilities that could lead to SQL injections, CSRF attacks, XSS attacks, or DDoS attacks. This capability assists defenders in identifying vulnerabilities and understanding various defence mechanisms.

While ChatGPT offers numerous benefits, including the ability to review and suggest improvements to code snippets, it also poses risks if used maliciously. For instance, malicious actors could exploit ChatGPT's functionality to enhance the effectiveness or obfuscation of malware. Therefore, careful consideration must be given to the potential dual-use nature of AI technologies like ChatGPT.

3) Code debugging

While GPT is not explicitly tailored for code debugging, developers have found practical utility in utilizing it for identifying code errors and vulnerabilities. The model can generate natural language explanations of code errors and propose potential solutions [47]. With adequate training data, it is anticipated that GPT will develop a deeper understanding of programming concepts and syntax, akin to experienced coders. In the future, ChatGPT is poised to advance further, enabling it to analyse code structure and identify logical loopholes in programs to mitigate security vulnerabilities [17].

Its capabilities extend to proficiently identifying significant coding mistakes, straightforwardly explaining intricate technical ideas, and aiding in the creation of scripts and robust code, among various other functions [47].

4) Cyberattack investigation and incidence response

ChatGPT offers extensive utility in analysing security incidents by processing large volumes of logs and text data to detect patterns and anomalies associated with cyberattacks. It rapidly generates natural language summaries of its findings, aiding cybersecurity experts and forensic analysts in comprehending the attack's scope, timeline, and characteristics for swift remediation. Utilizing ChatGPT as a powerful tool can greatly enhance threat detection and response within cybersecurity, especially in the detection of anomalies in network traffic and the analysis of user behaviour for identity and access management purposes.

Key applications of ChatGPT in cybersecurity incidents include:

Identifying Suspicious Activities in Logs: ChatGPT assists in reviewing log activity to identify potentially malicious actions or anomalies [49].

Summarizing Security Reports: It aids in summarizing breach reports, enabling analysts to understand attack methodologies and prevent future occurrences [17], [18].

Deciphering Attacker Code: Analysts can upload attacker code to ChatGPT for explanations of the steps taken and executed payloads, enhancing comprehension of attack techniques [17], [18]s.

Predicting Attack Paths: By analysing past cyberattacks and employing techniques, ChatGPT can predict future probable attack paths, enhancing proactive threat mitigation efforts [18].

ChatGPT contributes to the earlier detection of malicious activity and provides explainability of attacks as they unfold, offering valuable support as a Threat Detector and Security Advisor [51].

5) Configuration automation

Correct configurations are essential for the efficient functioning of software and cybersecurity tools and systems. Prompt engineering, involving the creation of carefully crafted instructions, enables ChatGPT to be trained to automatically configure various components such as servers, firewalls, intrusion prevention systems, and other cybersecurity tools [17].

By leveraging prompt engineering techniques, ChatGPT can generate scripts tailored to automate configurations effectively and securely. These scripts streamline the configuration process, ensuring adherence to security best practices while optimizing efficiency. As a result, ChatGPT becomes a valuable asset in the automation of configuration tasks, contributing to enhanced cybersecurity posture and operational efficiency within organizations.

6) AI-generated Phishing email detection

The "ChatGPT Detector" technology currently exists and is poised to evolve alongside ChatGPT itself. Ideally, IT infrastructure would incorporate AI detection software capable of automatically screening and flagging emails generated by AI [12].

On a positive note, ChatGPT has the potential to fulfil a crucial role, particularly in identifying phishing attempts [46]. Organizations could instil a practice among their employees to utilize ChatGPT to discern whether unfamiliar content is phishing or potentially generated with malicious intent [31]. This proactive approach can significantly enhance cybersecurity awareness and defence mechanisms within organizations [41].

7) Bridging the skills gap in cybersecurity

ChatGPT offers valuable assistance to cybersecurity teams facing challenges related to skills shortages by supporting tasks such as vulnerability discovery, forensic analysis, and report generation. By leveraging ChatGPT's capabilities, security teams can automate various processes, such as analysing extensive log files and generating executive reports, enabling them to concentrate on tasks that demand human analysis and expertise [32].

Moreover, the integration of AI in cybersecurity operations has the potential to mitigate the cybersecurity skills gap. Even junior personnel with limited cybersecurity experience can leverage AI-powered tools like ChatGPT to access answers and knowledge almost instantaneously. This democratization of information enables individuals across skill levels to contribute meaningfully to cybersecurity efforts, thereby bolstering overall resilience against cyber threats.

Furthermore, ChatGPT can offer valuable assistance to junior security workers by aiding in the communication of issues and enhancing their understanding of the context of their tasks. It can also support under-resourced teams by facilitating the curation of the latest threats and identification of internal vulnerabilities [32].

8) Security Research

ChatGPT serves as a rich source of insights, streamlining research and problem-solving tasks by granting users access to the vast corpus of the public internet with just one set of instructions. This resource empowers cybersecurity professionals to swiftly access information, search for answers, brainstorm ideas, and take proactive measures to detect and protect against threats more efficiently [32].

Moreover, ChatGPT facilitates the rapid acquisition of knowledge and understanding of new terms, processes, technologies, and methodologies relevant to cybersecurity. It provides immediate, concise and sometimes accurate answers to security-related queries, thereby shortening the time required for research and learning. This capability empowers cybersecurity professionals to stay abreast of the evolving threat landscape and adapt to emerging technologies and techniques efficiently.

ChatGPT's capabilities extend beyond individual cybersecurity efforts, fostering innovation and collaboration among researchers, practitioners, academia, and businesses in the cybersecurity industry. By harnessing the power of ChatGPT, stakeholders can drive advancements in cybersecurity practices, technologies, and strategies.

6. Conclusion and Future Scope

The research on the role of ChatGPT in cybercrime and cybersecurity through blog mining has yielded valuable insights into the potential impacts of this technology. Through the systematic analysis of publicly available blogs, we have elucidated the multifaceted capabilities of ChatGPT, its potential contributions to cybercrime, and its role in fortifying cybersecurity defences.

Our findings underscore the transformative potential of ChatGPT in various domains, from content generation to code development, customer service enhancement, and educational assistance. However, alongside its benefits, ChatGPT also presents significant challenges, particularly in its susceptibility to exploitation by malicious actors for cybercriminal activities. From the creation of sophisticated phishing attempts to the development of malware and other cyber threats, ChatGPT's capabilities can be harnessed for nefarious purposes.

In the realm of cybersecurity, our analysis reveals both opportunities and risks associated with ChatGPT. On one hand, it offers the promise of more efficient threat detection, enhanced incident response, and improved communication within security teams. On the other hand, its use in cybercrime underscores the need for robust defence mechanisms, including advanced threat detection systems, user education programs, and policy frameworks to mitigate potential risks.

Moving forward, stakeholders in the cybersecurity community must remain vigilant and proactive in addressing the evolving threat landscape shaped by technologies like ChatGPT. Collaboration between researchers, industry professionals, policymakers, and technology developers will be essential in developing effective strategies to harness the benefits of ChatGPT while minimizing its potential for misuse.

In conclusion, while ChatGPT holds immense potential for innovation and advancement across various domains, including cybersecurity, its responsible and ethical use is paramount. By understanding its capabilities, vulnerabilities, and implications, we can work towards harnessing the transformative power of ChatGPT for the collective benefit of society while safeguarding against potential risks and threats.

Data Availability

All included blog websites are publicly available.

Conflict of Interest

None

Funding Source

None

Authors' Contributions

The entirety of the work was independently conducted by the primary author, who served as the sole contributor to the research endeavour.

References

[1] P. V. Falade, "Decoding the Threat Landscape : ChatGPT , FraudGPT , and WormGPT in Social Engineering Attacks," *Int. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, no. October, 2023, doi: 10.32628/CSEIT2390533.

[2] S. Addington, "ChatGPT: Cyber Security Threats and Countermeasures," pp. 1–12, 2023.

[3] A. Qammar, H. Wang, J. Ding, A. Naouri, M. Daneshmand, and H. Ning, "Chatbots to ChatGPT in a Cybersecurity Space: Evolution, Vulnerabilities, Attacks, Challenges, and Future Recommendations," *J. Latex Cl. Files*, vol. 14, no. 8, pp. 1–17, 2023.

[4] P. V. Falade and P. O. Momoh, "Evaluating the Permissions of Monitoring Mobile Applications for Remote Employees : Analysing the Impact on Employer Trust and Employee Privacy Concerns," *Int. J. Sci. Res. Comput. Sci. Eng.*, vol. 12, no. February, pp. 42–52, 2024.

[5] G. Sebastian, "Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk ? An Exploratory Study," *Int. J. Secur. Priv. Pervasive Comput.*, vol. 15, no. 1, pp. 1–11, 2023, doi: 10.4018/IJSPPC.320225.

[6] M. Alawida, S. Mejri, A. Mehmood, B. Chikhaoui, and O. I. Abiodun, "A Comprehensive Study of ChatGPT : Advancements , Limitations , and Ethical Considerations in Natural Language Processing and Cybersecurity," *Inf. J.*, 2023.

[7] M. M. Mijwil, M. Aljanabi, and A. H. Ali, "ChatGPT : Exploring the Role of Cybersecurity in the Protection of Medical Information Intro to ChatGPT," *Mesopotamian J. Cybersecurity*, vol. 2023, pp. 18–21, 2023.

[8] P. Wagenseil, "Security risks of ChatGPT and other AI text generators," 2023.

[9] J. Cook, "How cybercriminals can use ChatGPT and Bard to attack businesses," 2023.

[10] Guru, “Beware That Hackers Using ChatGPT to Develop Powerful Hacking Tools,” 2023.
[11] K. Chakraborty, “ChatGPT at Risk: The Latest AI Package Hallucination Cyberattack,” 2023.
[12] J. Chilton, “The New Risks ChatGPT Poses to Cybersecurity,” 2023.
[13] C. Blackmon, “Cyber Criminals Are Using ChatGPT to Fool Victims,” 2023.
[14] L. Constantin, “Study shows attackers can use ChatGPT to significantly enhance phishing and BEC scams,” 2023.
[15] T. S. Dutta, “How Hackers Abusing ChatGPT Features For Their Cybercriminal Activities – Bypass Censorship,” 2023.
[16] C. Glover, “OpenAI’s ChatGPT safeguards ‘trivial to bypass’ for criminals, Europol says,” 2023.
[17] E. Maor, “ChatGPT’s cybersecurity implications: The good, the bad and the ugly,” 2023.
[18] TheHackerNews, “Offensive and Defensive AI: Let’s Chat(GPT) About It,” 2023.
[19] S. Sjouwerman, “The Implications Of ChatGPT On Cybercrime,” 2023.
[20] M. Hill, “Foreign states already using ChatGPT maliciously, UK IT leaders believe,” 2023.
[21] L. Bertuzzi, “Europol warns against potential criminal uses for ChatGPT and the likes,” 2023.
[22] G. Hibberd, “How ChatGPT is Changing Our World,” 2023.
[23] J. Burt, “Cybercrooks are telling ChatGPT to create malicious code,” 2023.
[24] N. C. Hughes, “Dark side of ChatGPT: it’s aiding cybercrime,” 2023.
[25] T. Keary, “Analysts share 8 ChatGPT security predictions for 2023,” 2023.
[26] C. Wisniewski, “Three Cybercrime Predictions In The Age Of ChatGPT,” 2023.
[27] D. B. Johnson, “Cybercriminals are already using ChatGPT to own you,” 2023.
[28] K. ALSPACH, “ChatGPT Is A ‘Powerful’ Tool For Cybercrime: Recorded Future,” 2023.
[29] D. Pandey, “ChatGPT’s powerful language model poses a threat to cybersecurity: Report,” 2023.
[30] S. Wadhwani, “Cybercriminals are Discussing How to Bypass ChatGPT Safeguards,” 2023.
[31] J. Perez-Etchegoyen, “The risk and reward of ChatGPT in cybersecurity,” 2023.
[32] R. Lariar, “ChatGPT is bringing advancements and challenges for cybersecurity,” 2023.
[33] NLTIMES, “AI tools like ChatGPT increasingly used by cybercriminals for phishing, experts warn,” 2023.
[34] J. Davidson, “How cyber criminals use ChatGPT to make better scams,” 2023.
[35] C. Lim and D. Ng, “As ChatGPT takes the world by storm, professionals call for regulations and defences against cybercrime,” 2023.
[36] T. Keary, “How ChatGPT can turn anyone into a ransomware and malware threat actor,” 2023.
[37] N. A. Malik, “ChatGPT — A New Dimension of Cyber Crime,” 2023.
[38] T. S. Dutta, “Hackers are Creating ChatGPT Clones to Launch Malware and Phishing Attacks,” 2023.
[39] D. GOODIN, “ChatGPT is enabling script kiddies to write functional malware,” 2023.
[40] P. Muncaster, “NCSC Calms Fears Over ChatGPT Threat,” 2023.
[41] A. Parwez, “Tech leaders’ perspectives on implementing ChatGPT and the looming cybersecurity,” 2023.
[42] S. Adlam, “ChatGPT has become a New tool for Cybercriminals in Social Engineering,” 2023.
[43] S. Sabin, “Hackers could get help from the new AI chatbot,” 2023.
[44] StudyFinds, “ChatGPT could help hackers launch devastating cyberattacks, experiments reveal,” 2023.
[45] iconJessica L. Hardcastle, “Russian criminals can’t wait to hop over OpenAI’s fence, use ChatGPT for evil,” 2023.
[46] Juan, “The risk and reward of ChatGPT in cybersecurity,” 2023.
[47] A. Scroxton, “Should we be worried about malicious use of AI language models?,” 2023.
[48] J. Banura, “What Cybersecurity Dangers Lurk Behind ChatGPT,” 2023.
[49] M. Ahmar, “Is chatGPT a threat to cybersecurity?,” 2023.
[50] A. MULGREW, “ChatGPT and cybersecurity: With great power comes great responsibility,” 2023.
[51] M. Hill, “Skyhawk adds ChatGPT functions to enhance cloud threat detection, incident discovery,” 2023.
[52] TheTimesOfIdia, “Similarly, on December 21, 2022, a threat actor dubbed USDoD,” 2023.
[53] V. Gain, “ChatGPT-4 can potentially make hackers’ lives easier, research finds,” 2023.
[54] H. Ravichandran, “How AI Is Disrupting And Transforming The Cybersecurity Landscape,” 2023.
[55] T. Brewster, “Armed With ChatGPT, Cybercriminals Build Malware And Plot Fake Girl Bots,” 2023.
[56] O. Powell, “Cybercriminals are using ChatGPT to create malware,” 2023.
[57] A. Mascellino, “ChatGPT Used to Develop New Malicious Tools,” 2023.
[58] S. Sabin, “Hackers are already abusing ChatGPT to write malware,” 2023.
[59] M. Marcelline, “Cybercriminals Using ChatGPT to Build Hacking Tools, Write Code,” 2023.
[60] E. GROLL, “ChatGPT shows promise of using AI to write malware,” 2023.
[61] S. R. Sunilkumar, “Cybercriminals using ChatGPT AI bot to develop malicious tools?,” 2023.
[62] Isb. Team, “How ChatGPT Can Help the Cybersecurity Sector?,” 2022.
[63] P. V. Falade, “Analysis of 419 Scams : The Trends and New Variants in Emerging Types,” *Int. J. Sci. Res. Comput. Sci. Eng.*, vol. 11, no. 5, pp. 60–74, 2023.

AUTHOR PROFILE

Polra Victor Falade- holds a B.Tech in Computer Science with a specialization in Cyber Security, which she earned from the Federal University of Technology Minna, Niger State, Nigeria in 2016. Subsequently, pursued an MSc in Information Security from the University of Surrey, UK, graduating in 2021. These educational experiences have equipped her with a comprehensive understanding of cybersecurity principles and best practices. Currently, she is serving as an Assistant Lecturer in the Department of Cyber Security at the Nigerian Defence Academy (NDA) in Kaduna, Nigeria. In this role, she has been actively involved in educating future cybersecurity professionals, fostering a culture of cybersecurity awareness, and conducting research in the field. Her commitment to academic excellence is reflected in her continuous pursuit of knowledge and dedication to her students. Furthermore, she is a professional member of the Cyber Security Expert Association of Nigeria (CSEAN), which has provided her with valuable networking opportunities and a platform to stay updated with the latest developments in the cybersecurity domain. Also, a member of Internet Society, Nigeria Chapter. Her primary passion lies in research and academic writing, particularly in the areas of Information Security, AI Security, Privacy and cybersecurity-related research.