Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 5, No. 1, March, 2024

Boundaries and Future Trends of ChatGPT Based on AI and Security Perspectives

Albandari Alsumayt [1*], Zeyad M. Alfawaer [1], Nahla El-Haggar [1], Majid Alshammari [2], Fatemah H. Alghamedy [1], Sumayh S. Aljameel [3], Dina A. Alabbad [4], May Issa Aldossary [5]

[1] *Department of Computer Science, Applied College, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.*

[2] *Department of Information Technology, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia.*

[3] *Saudi Aramco Cybersecurity Chair, Computer Science Department, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.*

[4] *Computer Engineering Department, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.*

[5] *Saudi Aramco Cybersecurity Chair, Computer Information Systems Department, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.*

Received 01 September 2023; Revised 18 February 2024; Accepted 24 February 2024; Published 01 March 2024

Abstract

In decades, technology and artificial intelligence have significantly impacted aspects of life. One noteworthy development is ChatGPT, an AI-based model that has created a revolution and attracted attention from researchers, academia, and organizations in a short period of time. Experts predict that ChatGPT will continue advancing, bringing about a leap in artificial intelligence. It is believed that this technology holds the potential to address cybersecurity concerns, protect against threats and attacks, and overcome challenges associated with our increasing reliance on technology and the internet. This technology may change our lives in productive and helpful ways, from the interaction with other AI technologies to the potential for enhanced personalization and customization to the continuing improvement of language model performance. While these new developments have the potential to enhance our lives, it is our responsibility as a society to thoroughly examine and confront the ethical and societal impacts. This research delves into the state of ChatGPT and its developments in the fields of artificial intelligence and security. It also explores the challenges faced by ChatGPT regarding privacy, data security, and potential misuse. Furthermore, it highlights emerging trends that could influence the direction of ChatGPT's progress. This paper also offers insights into the implications of using ChatGPT in security contexts. Provides recommendations for addressing these issues. The goal is to leverage the capabilities of AI-powered conversational systems while mitigating any risks.

Keywords: ChatGPT; Artificial Intelligence; Security; Privacy; Cyber Security; Attacks; LLMs.

1. Introduction

Open AI, founded in 2015 by Elon Musk, Sam Altman, and others, is dedicated to developing artificial general intelligence (AGI) for the betterment of humanity. Their remarkable journey includes the creation of groundbreaking AI

* Corresponding author: afaalsumayt@iau.edu.sa

http://dx.doi.org/10.28991/HIJ-2024-05-01-010

models like GPT-2, GPT-3, and the latest innovation, ChatGPT. Building upon the success of GPT-3, Open AI pursued further research and development to introduce ChatGPT, based on the advanced GPT-4 architecture. ChatGPT outshines GPT-3 in conversation-based tasks, elevating contextual understanding, response generation, and overall coherence to new heights. Open AI's primary objective with ChatGPT is to achieve superior contextual comprehension, generate more coherent responses, and enhance overall coherence [1].

In the realm of Natural Language Processing, a significant breakthrough has been achieved through the development of large language models (LLMs). These remarkable artificial intelligence models are specifically designed to comprehend and analyze human language. According to the Yang et al. [2] study, LLMs are characterized by four key features that set them apart: a profound understanding of natural language text, the ability to generate text that resembles human language, contextual awareness, and exceptional problem-solving and decision-making capabilities. These features have propelled LLMs to the forefront of advancements in Natural Language Processing, paving the way for groundbreaking developments in the field.

In 2023, a multitude of LLMs emerged, captivating audiences and gaining widespread acclaim. Notable examples of these advanced language models include OpenAI's ChatGPT [3] and Meta AI's LLaMA [4]. The sheer popularity of ChatGPT is evident, as it boasts an impressive millions of users. LLMs have now expanded their horizons, offering a diverse range of versatile applications across various domains. They not only provide technical support in language processing-related fields like search engines [3, 5, 6], customer support [7], and translation [8, 9], but also prove valuable in general scenarios such as code generation [10], healthcare [11], finance [12], and education [13]. This remarkable adaptability highlights their potential to streamline language-related tasks across industries and contexts, making them invaluable tools in today's world.

Leveraging language modeling techniques, ChatGPT undergoes extensive pre-training on a diverse corpus of text data, including books, articles, and websites [14]. This pretraining equips ChatGPT with the ability to generate coherent and realistic responses during conversations, enabling it to learn intricate patterns and relationships between words. The multilingual capabilities of ChatGPT make it exceptionally versatile, as it can seamlessly integrate into global applications and cater to a wide range of users. Translation, sentiment analysis, and multilingual content creation are just a few examples of its indispensable applications [15].

ChatGPT consistently delivers grammatically correct, coherent, and contextually accurate text, making it a valuable tool for various purposes such as content writing, summarization, and rewriting [16]. Its contextual understanding empowers ChatGPT to grasp the nuances of text-based conversations, resulting in more natural and engaging interactions with users. By leveraging its understanding of sentences and phrases, ChatGPT generates relevant and coherent responses, ensuring smoother communication [17]. Refining or breaking down prompts using prompt engineering techniques allows users to guide ChatGPT towards desired information. This approach significantly enhances the quality and effectiveness of ChatGPT conversations.

ChatGPT, with its remarkable ability to generate convincing responses, has become a powerful tool. However, this capability also opens the door for malicious actors to exploit its potential for spreading disinformation, launching phishing attacks, and impersonating individuals [17, 18]. Consequently, it is crucial to continuously monitor and assess ChatGPT's security vulnerabilities while developing appropriate mitigation measures.

The risks associated with ChatGPT's exploitation are far-reaching, with potential consequences ranging from financial losses and data breaches to privacy violations. Of particular concern is the ease with which highly convincing phishing attacks can be generated using ChatGPT, posing a significant security risk. Attackers can manipulate conversations to their advantage, further exacerbating the potential damage [17]. To safeguard against these risks, it is essential to implement robust security measures that address the diverse challenges posed by ChatGPT. By doing so, we can protect individuals and organizations from the adverse effects that stem from its misuse.

The growing popularity of large language models (LLMs) within the security community has sparked numerous research papers highlighting their applications in security and privacy. These papers encompass a range of perspectives, including those that emphasize the positive impact of LLMs on security, explore potential risks to security, and delve into discussions on security vulnerabilities inherent in LLMs [19]. LLMs have proven to be instrumental in bolstering code security, data security, and privacy within the security community. Their positive influence is evident in their ability to enhance these aspects. However, it is important to acknowledge that LLMs can also be utilized for offensive purposes against security and privacy. These offensive applications encompass a wide array of attacks, including hardware-level attacks, OS-level attacks, software-level attacks, network-level attacks, and user-level attacks.

The exploration of LLMs in the realm of security highlights both their potential for positive contributions and the need for vigilance to mitigate potential risks. By understanding the multifaceted nature of LLMs in the context of security, the community can work towards harnessing their benefits while proactively addressing any associated challenges.

However, as with any AI application, ChatGPT raises ethical concerns and risks of misuse. Its powerful text reasoning and generation capabilities have led students to find it incredibly helpful for completing their homework. Initially used to explain difficult concepts or rephrase project reports, it quickly became a tool for writing entire assignments [20]. This misuse caught the attention of teachers and schools, leading to its identification as plagiarism. Another concern revolves around the copyright issues stemming from ChatGPT. As more people use it to generate original-like text content without proper citation, the question of copyright ownership for the content created by ChatGPT becomes a serious issue. With no one taking responsibility for the accuracy and correctness of the generated content, regulating the copyright of machine-generated visual and textual content becomes imperative.

While ChatGPT incorporates privacy protection mechanisms, such as blocking access to personal data, there is still a risk of potential data leakage. Malicious attacks, like jailbreaking, could exploit its powerful generation abilities to infer information from personal data or even launch attacks on other AI models [21]. It is widely recognized that, despite the numerous advantages ChatGPT offers, the potential security, privacy, and ethical problems associated with it cannot be ignored [22]. Several research works have been proposed to address these problems, although only a few have attempted to consolidate and summarize them. To advance these solutions and pave the way for future work, it is crucial to compile these efforts, providing a comprehensive comparison and analysis to improve upon existing approaches and explore new directions. The ethical implications of ChatGPT, such as generating malicious, offensive, and biased content, are among the primary concerns surrounding this technology. In this article, we aim to contribute to the ongoing discussion on the potential improvement of security measures when utilizing ChatGPT. Our goal is to gain a deeper understanding of how this technology can be leveraged to enhance security.

The structure of this paper is as follows: Section 2 explores related works and identifies the open problems in the proposed research solution. In Section 3, we delve into the role and significance of security in the utilization of ChatGPT, as well as the emerging attack trends. Section 4 presents several security mechanisms that can be implemented to enhance the level of security in ChatGPT. Additionally, Section 5 provides insights into the future evolution of ChatGPT based on artificial intelligence. Section 6 discusses the possible future trends in ChatGPT based on AI and security. Section 7 explains the ethical implications and recommendations of ChatGPT based on both AI and security. Finally, we conclude the paper by outlining future research directions about ChatGPT. Figure 1 shows the methodology and scope of this study.

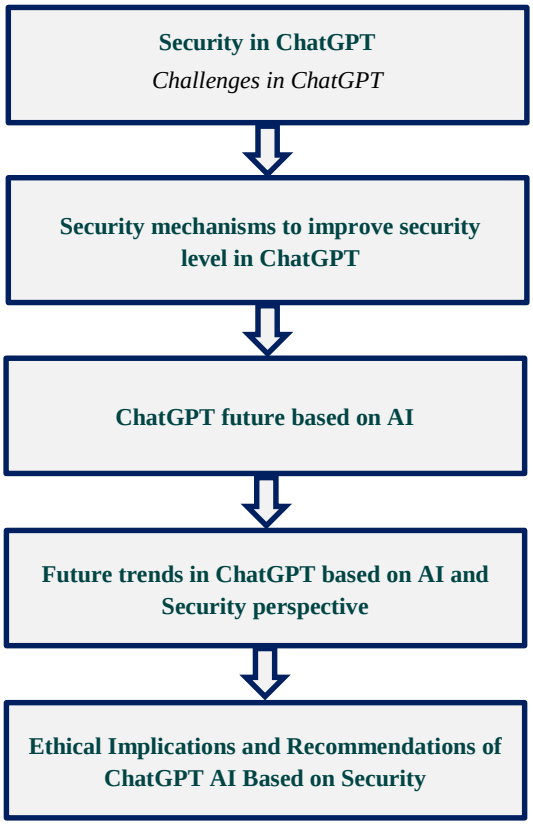

Figure 1. Methodology and scope of the study

2. Related Works

Khoury et al. [23] experimented to assess the safety of code generated by GPT-3.5 using ChatGPT. They instructed ChatGPT to generate 21 programs in five different programming languages, specifically chosen to highlight risks

associated with specific vulnerabilities. Notably, no security features were requested during the code generation process. The findings revealed that although 80% of the generated code was executable, it did not meet the minimum standards for secure coding. Less than 25% of the generated code was considered secure against a specific vulnerability, and this percentage could be even lower if additional vulnerabilities were included. However, with the assistance of expert interactions, ChatGPT was able to rectify approximately 45% of the insecure code.

Renaud et al. [24] discussed how advanced technologies like ChatGPT introduce new methods for cybercriminals to achieve their objectives. They highlighted that ChatGPT could comprehend the security design of targeted systems, and its capacity to generate AI-driven languages enhanced the quality of deceptive communications. The authors suggested that traditional security policies and best-practice approaches may prove ineffective in the era of ChatGPT. They proposed several methods to enhance security in response to this new attack style. For instance, incorporating ChatGPT and other AI-generative models with mail servers could help detect whether suspicious emails are AI-generated. Additionally, they emphasized the importance of knowledge-based preparedness through awareness training to detect and respond to emerging threats. For further details on ChatGPT-related works, please refer to Table 1, while Table 2 provides an overview of ChatGPT security risks.

Table 1. ChatGPT related works

Paper Title	Summary
Beyond the Safeguards: Exploring the Security Risks of ChatGPT [25]	In this paper the authors explored six security risks: information gathering, malicious text writing, malicious code generation, dis-closing personal information, fraudulent services, and providing unethical content. In this paper the authors selected cases with examples of real interactions with ChatGPT to demonstrate these security risks in practice by writing the prompt and the response.
Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk? [26]	In this paper, the authors talked about cyber risks associated with the use of ChatGPT: Social engineering attacks, Malware threats, Phishing attacks, Identity theft, Data leakage. The paper included surveys conducted on cybersecurity attacks associated with ChatGPT. It also stated methods to minimize these cyber threads.
How Secure is Code Generated by ChatGPT? [23]	In this study, they performed an experiment to address the safety of generated code by GPT-3.5. They asked ChatGPT to generate 21 programs, in 5 different programming languages, each of which is chosen to highlight risks of a specific vulnerability. Besides that, they did not ask ChatGPT to include any security features. Even though 80% of the generated codes are executable, codes indicate that ChatGPT is not able to generate codes that meet the minimum standards requirements of secure coding in which less than 25% of the generated codes are considered a secure code against a specific vulnerability and could be less if they include more vulnerabilities. However, ChatGPT can fix about 45% of the insecure code with the help of expert interactions.
Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information [27]	In this paper, the Author talked about the privacy risks and concerns associated with ChatGPT, such as: data poisoning, data leakage and sharing of sensitive information. in this paper the authors proposed some techniques that can increase the privacy protection. They also evaluated every technique and measured the impact on the performance of ChatGPT.
From ChatGPT to HackGPT Meeting the Cybersecurity Threat of Generative AI [24]	The authors in this paper discussed how "smarter" technology such as ChatGPT introduces new methods for cybercriminals to attain their targets. This is because of several reasons such as ChatGPT can understand the security design of targeted systems. In addition, the capacity of ChatGPT for producing AI-driven languages boosts the quality of fake communications. The traditional security policies such as best practice approaches could be useless in the era of ChatGPT. They propose several methods that may help to raise the security level under the new attack style such as incorporating ChatGPT and other AI generative with mail servers to detect whether the suspect emails are AI-generated or not may support the security. In addition, awareness training should be knowledge-based preparedness to detect new threats.
Potential Risks of ChatGPT: Implications for Counterterrorism and International Security [28]	This paper aims to study the implications of ChatGPT tools for the field of international stability and security. The study highlighted four key points: the implications of artificial intelligence for future threats and its effect on international security; the impact of ChatGPT on cyberterrorism and artificial intelligence and how it creates new opportunities and approaches for the field of cyberterrorism; the dangers of fragmented information for violence and disruption operations; the use of psychological warfare against targets by misusing ChatGPT and crating fake news and terrorist activities.

Table 2. ChatGPT Security Risks

ChatGPT Security Risks	Paper title
{Providing Unethical Content, Private Data disclosure, Information Gathering, Malicious code generation, Fraudulent services, malicious text writing}	Beyond the Safeguards: Exploring the Security Risks of ChatGPT [25]
{Social engineering attacks, Malware threats, Phishing attacks, Identity theft, Data leakage}	Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk? [26]
{Unintended Sharing of Sensitive Information, Data Leakage, Adversarial Attacks, Model Extraction, Data Poisoning}	Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing Us-er Information [27]

3. Security in ChatGPT

Large Language Models (LLMs) have raised concerns regarding various implications, including risks associated with private data disclosure, the generation of offensive content, and the potential for generating malicious code.

Research highlighted by Derner & Batistič [25] suggests that LLM models like ChatGPT are susceptible to numerous vulnerabilities, such as data leakage, code injection, unauthorized code execution, training data poisoning, insufficient access controls, improper error handling, overreliance on LLM-generated content, and inadequate sandboxing [16]. These vulnerabilities are illustrated in Figure 2, which provides an overview of the main security concerns in ChatGPT.

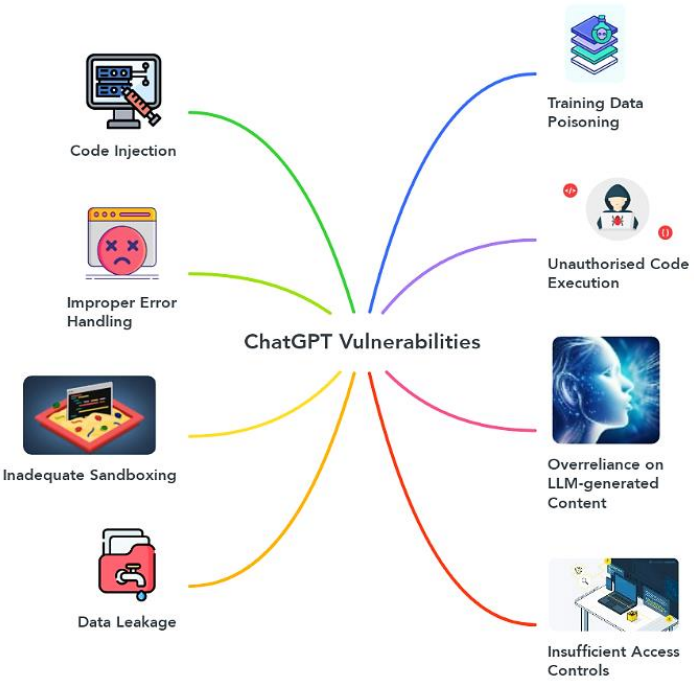

Figure 2. Vulnerabilities in ChatGPT

ChatGPT, like other large language models (LLMs), is not immune to security vulnerabilities. These vulnerabilities can impact its performance and pose risks to users. The following vulnerabilities have been identified:

- Data leakage: ChatGPT may inadvertently disclose sensitive information, proprietary algorithms, or sensitive details in its responses. Although incidents of data breaches have been quickly addressed and had minimal impact, they highlight potential risks for chatbots and users in the future [29].
- Code injection: Attackers can modify chatbot answers using invisible single-pixel mark-down images, enabling them to extract sensitive user data. This type of attack can persist and affect future answers, even without exploiting specific vulnerabilities. Additionally, ChatGPT's accessibility empowers novice hackers to generate malicious code without deep technical knowledge [23].
- Unauthorized code execution: Exploiting the natural language prompts, attackers can execute malicious code, actions, or commands on the system by leveraging the capabilities of LLMs [30].
- Training data poisoning: Deep learning models rely on massive training datasets collected from web crawling. However, trust in this data is increasingly threatened by data poisoning attacks, where intentionally malicious information compromises the training data. Countermeasures are being explored to make falsifying records more challenging [31].
- Insufficient Access Controls: Poorly implemented access controls or authentication mechanisms can enable unauthorized users to interact with LLMs and exploit vulnerabilities [32].
- Improper error handling: Inadequate error handling can result in the disclosure of error messages or debugging information, which may expose sensitive information, system de-tails, or potential attack vectors [33].
- Overreliance on LLM-generated Content: Excessive reliance on LLM-generated content without human oversight can have detrimental consequences. Human supervision is crucial to ensure the quality, accuracy, and appropriateness of the generated content [2].
- Inadequate sandboxing: Inadequate sandboxing can lead to security risks and compromises. Implementing robust sandboxing mechanisms is essential to prevent unauthorized access and malicious activities [34].

In summary, ChatGPT exhibits various security vulnerabilities that need to be considered. Table 3 provides an overview of the security considerations associated with using ChatGPT.

Table 3. Security Considerations in ChatGPT

Security Factor	Explain	Consideration
Data privacy and security	The raise usage of AI in data analysis and processing leads to make data security more prevalent.	Need to consider the sensitivity of data and protect the security and privacy.
Explain ability and transparency	ChatGPT is a complex model and not easy to explain or understand.	It is essential to ensure transparency in areas where decision is vital and whole data as well.
Misinformation	The ability of ChatGPT to interact with people such as humans increases the impact of AI systems on human autonomy.	Individuals need to maintain control over their selections and actions.
Autonomy	The ability of ChatGPT to interact with people such as humans increases the impact of AI systems on human autonomy.	Individuals need to maintain control over their selections and actions.
Bias and discrimination	Large datasets are trained that might cause biases.	The model may learn these biases and generate responses that could be offensive.
Misuse and abuse	ChatGPT can be used for malicious goals. Such as, generate fake news, and impersonation.	Ensure that ChatGPT is used ethically. Some procedures can help to protect people such as produce safeguards and filter contents.
Privacy and security	ChatGPT can be used in sensitive datasets such as medical reports, private messages, and financial records.	These datasets must be kept securely, and only legitimate users can access them.
Fairness	While ChatGPT is trained over massive data from the Internet, it might propagate bias and absorb in the training data.	Output might be reinforcing stereotypes and need to improve rules to debias AI models and generate fair algorithms.
Accountability and responsibility	ChatGPT becomes more powerful	Identify the person who is responsible for making decisions and confirm actions. Questions need to be considered: • Who is accountable for the bad consequences of using this technology? • Who owns the data? • Who is responsible for results that generated by ChatGPT?

4. Security Mechanisms to Improve Security Level in ChatGPT

Ensuring user privacy and minimizing security risks in the realm of large language models such as ChatGPT is a complex undertaking that necessitates the implementation of various techniques. These techniques encompass differential privacy, secure multi-party computation, privacy-aware machine learning algorithms, adversarial training, robustness testing, rate-limiting, blocking automated queries, anonymization, and encryption methods [35]. By employing these methods, it becomes possible to guarantee that user data remains appropriately safeguarded against unauthorized access and misuse throughout both the model training and interaction phases.

When it comes to AI model training and testing, preserving data privacy assumes paramount significance, particularly in instances involving sensitive or confidential information [36, 37]. However, achieving comprehensive privacy preservation in AI necessitates the consideration of the four pillars of Privacy-Preserving Machine Learning (PPML): training data privacy, input privacy, output privacy, and model privacy. The first three pillars primarily focus on protecting the privacy of data creators, while the fourth pillar aims to safeguard the privacy of model creators. A taxonomy delineating privacy preservation techniques can be observed in Figure 3, and Table 4 offers a comparative analysis of these techniques [37].

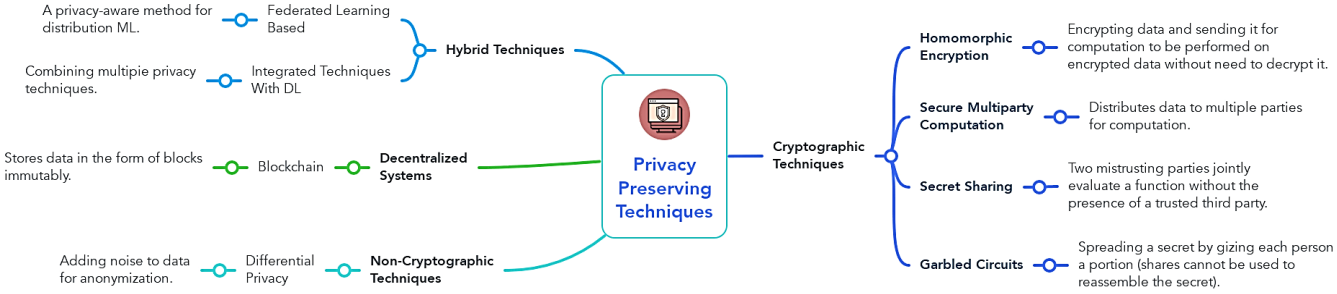

Figure 3. Taxonomy of privacy-preserving techniques

Table 4. Comparison between privacy preservation techniques

Type	Privacy preservation techniques	Description	Advantages	Disadvantages
Cryptographic techniques	Homomorphic encryption (HE)	Homomorphic encryption (HE) is a technique in which the linear models are transformed into encrypted models using Pillar methods. It involves the development of a collective learning protocol, which facilitates the exchange of classified time-series data within an organization. The encryption of data is performed by the data owner, and the decryption occurs after all computations are completed.	Secure and efficient cloud utilization and collaboration with third parties. It can also be utilized to obtain outsourced services for research and analysis while maintaining an imputation efficacy of over 0.99 AUC score through multiple optimizations.	Slow and requires either a programmed or dedicated client-server application for proper functioning.
	Secure Multi- party Computation (SMPC) [38]	Data can be computed by dividing it among various parties, who then apply the algorithm to their respective secure data without being aware of the other data. This approach helps maintain privacy and enables multiple parties to perform a function on their inputs while keeping them confidential. However, this technique may not be suitable for training large language models such as GPT-4, which usually involves a single dataset owned by a single entity.	Even if the input data is searched for an indefinite amount of time and resources, it will remain private as there are many parties involved who cannot be trusted and may have ulterior motives.	The computation process requires assumptions about the number of untrustworthy parties involved, which can result in higher costs for communication and computation, leading to decreased performance. The effect on usability will vary depending on how the implementation is executed.
Non cryptographic techniques	Differential Privacy Techniques [39, 40]	The concept of differential privacy ensures that privacy is maintained even when the adversary has extensive external knowledge. The approach involves adding sufficient noise to the outcome, such as the model produced from training, to conceal the contribution of any individual to that outcome. This technique is based on a theoretical foundation and aims to prevent the model from learning too much about any specific example in the training data by incorporating a regulated quantity of noise.	Differential privacy is a robust method that allows for modular design and study of privacy techniques due to its ability to be composed, withstand post-processing, and gracefully degrade when dealing with correlated data.	Differential privacy is more effective at adding noise to data than previous methods because it not only prevents linking but also prevents reconstruction. However, this method may result in a trade-off between privacy and utility, as the added noise can negatively impact the performance and accuracy of the model.
	Rate-Limiting and Blocking Automated Queries	Rate-limiting involves constraining the number of requests made by a user or service to the system within a designated time, it is serving as a protective measure that can be implemented is blocking automated queries to shield the system from automated attacks or misuse. Blocking automated queries is another technique used to prevent abuse by bots or automated scripts. These queries are typically made to extract data or perform actions that may negatively impact the application or violate its terms of service. By detecting and blocking such automated queries, the application can protect its resources and ensure fair usage for all users.	This approach provides defense against various forms of attacks, including DDoS attacks, credential stuffing, brute force attacks, and data scraping.	A limit on the number of requests a user or IP address can make within a specific timeframe.
	Adversarial Training, Robustness Testing [41, 42]	Highly effective method for protecting deep learning models from adversarial examples by reducing the malicious effect caused by adversarial attacks. Unlike other defense strategies, it focuses on improving the models themselves to enhance their intrinsic robustness.	The quality of the adversarial samples used during training is crucial in addressing issues like overfitting, generalization, and training efficiency. Increase the robustness of a model against adversarial attacks and help to improve its generalization.	Higher computational costs and intricacy, potentially impacting the overall performance usability.
	Privacy-Aware Machine Learning Algorithms (FL) [43]	(FL) aims to ensure privacy during the learning process by distributing data among various groups and companies, creating separate datasets. This approach preserves local privacy while enabling real-time continual learning and diverse data.	Data protection involves storing the training dataset on individual devices, which eliminates the necessity for a centralized data pool. This allows for real-time continual learning and ensures data diversity.	There are still challenges to overcome, such as attacks on robustness and the need for improved efficiency and effectiveness. Additionally, there may be computational overhead or a decrease in model accuracy in the current state of FL.
	Blockchain (Decentralized system) [27]	Blockchain is a method that ensures privacy and secures personal information using private key encryption and zero-knowledge proofs.	decentralization, immutability, transparency, and access control.	Complexity, publicly accessible blockchains, scalability issues, and vulnerability to data breaches
Hybrid privacy-preserving deep learning (HPPDL)	Anonymization and encryption techniques [27, 39]	These techniques are key to protecting data and privacy, ensure that data used in training and interaction is secure, and not vulnerable to unauthorized access or misuse. In AI, anonymization is commonly used to protect user data during model training, and techniques include data masking, pseudonymization, generalization, and differential privacy.	-Secure the data used during training and interaction, -Removing personally identifiable information from datasets to prevent linking the data back to the individual it originated from. -Preventing unauthorized access and misuse.	There is a risk of confidentiality breaches if the model produces results that reveal sensitive patterns in the data.
	FL+HE, FL+SMPC, FL+DP, FL+DL & BC [44-46]	The integration of FL with SMPC and DP offers sufficient security measures to comply with General Data Protection Regulation GDPR, which demands strict data security and protection. These technologies aim to overcome the hurdles imposed by GDPR and ensure the secure collection and utilization of large datasets.	This combination incorporates adequate security measures to fulfill data protection requirements Effectively protect privacy, reduce the cost of training ML models, and make use of diverse community-sourced data.	The time complexity increases. There is a trade-off between accuracy and efficiency.

5. ChatGPT Future Based on AI

Artificial Intelligence (AI) has made remarkable progress across various sectors, including cybersecurity. The emergence of sophisticated AI models like OpenAI's ChatGPT has brought about a significant shift in security operations. However, like any technology, AI advancements in cybersecurity carry both advantages and disadvantages.

ChatGPT, like other AI applications, encounters certain limitations and challenges in the realm of AI and security. While it offers transformative capabilities, there are ethical concerns and risks of misuse associated with its use. One notable challenge is the potential misuse of ChatGPT in academic settings. Due to its powerful text-generation abilities, students have found it helpful for completing homework assignments. However, this has led to instances of plagiarism, as students rely on ChatGPT to write entire assignments without proper attribution. Another concern revolves around the copyright implications of content generated by ChatGPT. As more individuals use ChatGPT to create original text content without proper citation, the issue of copyright ownership becomes significant. The lack of responsibility for the accuracy and correctness of the generated content raises questions about the regulation of machine-generated visual and textual content [47].

In terms of security, there are potential risks associated with the use of ChatGPT. Cybercriminals and fraudsters may exploit the technology for destructive purposes, such as creating scripts for dark web marketplaces. It is crucial to implement strict security measures, ensure responsible and ethical usage, and continuously monitor and update the model's capabilities to mitigate these risks [48]. The ethical, legal, and societal implications of harnessing ChatGPT in various AI and security applications are also worth considering. The reliance on ChatGPT for conversations raises concerns about the loss of genuine human connection and the potential detrimental effects on society. Additionally, AI models like ChatGPT can propagate inaccuracies or biases present in the training data, highlighting the need for continuous improvement in the training process to mitigate bias.

Addressing these limitations and challenges requires ongoing research and development. Efforts should focus on curating diverse and high-quality datasets, implementing bias mitigation techniques, and promoting responsible usage to ensure the ethical and secure deployment of ChatGPT in different applications. In summary, while ChatGPT offers transformative capabilities, it is essential to address the ethical concerns, security risks, and limitations associated with its use. By actively addressing these challenges, researchers and developers can pave the way for responsible and beneficial applications of ChatGPT in AI and security domains.

5.1. Pros of AI Advancements in Cybersecurity

- Enhanced Threat Detection: AI models like ChatGPT can be trained to identify patterns and anomalies in data that might indicate a cybersecurity threat. This capability can significantly enhance threat detection and response times. For instance, AI can analyze network traffic and identify unusual patterns that might suggest a potential cyberattack [49].
- Automation of Routine Tasks: AI can automate routine tasks, freeing up cybersecurity professionals to focus on more complex issues. For example, ChatGPT can be used to automate the generation of phishing emails for security awareness training or to automate responses to common security inquiries [50].
- Proactive Security Measures: AI can help in predicting and preventing cyber-attacks before they occur. By analyzing historical data, AI can identify patterns and predict future attacks, enabling organizations to take proactive security measures [51].

5.2. Cons of AI Advancements in Cybersecurity

- Dependence on Data Quality: The effectiveness of AI models like ChatGPT heavily depends on the quality of the training data. If the data is biased, incomplete, or inaccurate, the AI model may produce unreliable results, which can have serious implications in a cybersecurity context [52].
- Risk of AI-Powered Cyber Attacks: While AI can enhance cybersecurity, it can also be used by cybercriminals to carry out sophisticated attacks. For instance, ChatGPT could be used to generate convincing phishing emails or to automate the discovery of system vulnerabilities [53].
- Lack of Explain ability: AI models often suffer from a lack of explaining ability, meaning it can be difficult to understand why they made a particular decision. This can be problematic in a cybersecurity context, where understanding the reasoning behind threat detection can be crucial [54]. AI advancements like ChatGPT offer promising benefits for cybersecurity, but they also present new challenges that need to be addressed. As with any technology, it is crucial to understand and mitigate these risks to fully leverage the potential of AI in cybersecurity.

6. Future Trends in ChatGPT Based on AI and Security Perspective

The domain of Chatbot Generative Pretrained Transformer (ChatGPT) technology is expected to witness significant future developments, closely intertwined with advancements in artificial intelligence (AI) and cybersecurity. This section delves into the potential evolution of ChatGPT technology and its impact on the AI and security domains.

6.1. Investigation of Possible ChatGPT Tech Progress and Innovations

- Enhanced Language Understanding and Generation: Future iterations of ChatGPT are poised to make substantial advancements in language processing, leveraging transformer models and self-supervised learning techniques to enhance contextual comprehension [55]. The incorporation of multimodal data processing is expected to further augment its language processing capabilities, aligning with recent developments in this field [56].

- Autonomous Learning and Adaptation: Future versions of ChatGPT are projected to feature advanced autonomous learning, enabling real-time adaptation to user interactions through continual learning algorithms [57].
- Human-AI Collaborative Interfaces: The trajectory of ChatGPT's evolution points toward more seamless human-AI interactions, indicating its potential to become a more collaborative tool. Ongoing research in Natural Language Processing (NLP) and Human-Computer Interaction (HCI) suggests that AI systems, including ChatGPT, may become proficient in understanding complex human intentions [58].

6.2. Prediction of the Future Trends and Their Effect on AI and Security Domains

- Impact on AI Development: The advancements in ChatGPT are anticipated to lead to more context-aware, ethically aligned AI systems that can better understand diverse human contexts, thereby addressing current limitations in AI adaptability across various domains [59].
- Security Implications: The advancement of ChatGPT raises complex security implications, including its application in cybersecurity for threat detection and user authentication. However, this progress also underscores the need for advancements in AI ethics and security protocols to mitigate potential misuse [60].
- Privacy and Data Security: As ChatGPT becomes more integrated into daily activities, ensuring data privacy and security becomes imperative. Implementing privacy-preserving techniques such as federated learning and differential privacy is crucial for safeguarding user data while maintaining AI efficiency [61].

The future development of ChatGPT technology is expected to significantly impact the AI and cybersecurity fields, promising enhanced capabilities while underscoring the importance of addressing ethical and security challenges in AI development. A comprehensive strategy encompassing technological innovation, ethical considerations, and robust security measures is essential for realizing ChatGPT's full potential in the evolving digital era.

7. Ethical Implications and Recommendations of ChatGPT AI Based on Security

The discussion of ethical considerations is necessary due to the wide array of potential applications for ChatGPT, an emerging technology with numerous possible uses. The ethical issues surrounding the exploitation of ChatGPT primarily revolve around privacy, bias, and the potential for misuse, as excessive use of this powerful tool may raise concerns about data accumulation, collection methods, and storage practices. Data collection must adhere to ethical standards and governmental laws aimed at preventing privacy infringements. As such, the strategy employed in the development and use of ChatGPT AI chatbots is ethically oriented and has the potential to bring about meaningful societal and human advancements. Therefore, it is essential to critically analyze the trade-offs between the use of AI chatbots for convenience and the preservation of human communication as an art form, taking into account social dynamics and expertise [22, 62]. Various interpretations and recommendations are outlined below for consideration [63, 64].

7.1. Ethical Implications

The advancement of ChatGPT technology raises significant ethical concerns related to security, data privacy, potential misuse, and its impact on human communication and social skills.

- Security concerns: The realistic nature of ChatGPT conversations presents security risks, including the potential for social engineering, phishing hoaxes, and impersonation. Cybercriminals can exploit ChatGPT to deceive victims into clicking on malicious links, sending sensitive information through fake emails, and installing malware. Moreover, the tool's advanced impersonation capabilities enable the AI to masquerade as a victim's associate or family member, undermining trust.
- Privacy Concerns and Data Protection: ChatGPT AI systems can collect and process substantial amounts of personal user data, including conversation logs, internet history, and location information. This raises significant privacy concerns, as the use or disclosure of such data to third parties without consent can infringe upon users' privacy rights. To address this, there have been calls for the implementation of data protection governance and user boundaries to ensure greater accountability and transparency in the collection and utilization of data by ChatGPT AI systems [20, 65].
- Protection of the training data and models: While ChatGPT aims to produce human-like text, there is a risk of unintended outcomes, including misinterpretation and the generation of offensive or harmful content. Additionally, the performance of ChatGPT can be influenced by factors such as the quality of training data and model structure, with potential implications for spamming and financial information theft.
- The performance of ChatGPT: The performance of ChatGPT is contingent upon factors such as the quality of its training data and model structure. However, there is a need to exercise caution against the misuse of machine learning, which could lead to spamming, theft of financial information, and impersonation-based attacks with detrimental effects on businesses. Additionally, there are concerns about the potential misuse of ChatGPT for executing impersonation and social engineering techniques [66].

- The potential for misuse of ChatGPT: There are concerns that ChatGPT's AI chatbot may be abused for malicious purposes, including cyber-attacks, dissemination of false information, and fraudulent activities such as phishing. Furthermore, the technology's impact on human communication and social skills is an ethical consideration, with some studies suggesting potential negative effects on empathic concern and satisfaction with social support, particularly for individuals experiencing social isolation [67].
- The potential effects of ChatGPT AI chatbot on human Interaction and social skills: The potential influence of ChatGPT AI chatbot on human interaction and social skills is an ethical concern worth exploring. While some studies suggest that chatbots can alleviate social isolation among older adults [68], others raise alarms about the potential adverse effects of chatbot use on human communication. For instance, research by Tsai & Chuan [69] has revealed that engaging with chatbots correlates with reduced empathic concern and decreased satisfaction with social support, especially for individuals already facing social isolation. This calls for greater attention to the societal and psychological implications of AI systems, along with increased resources dedicated to nurturing social skills and support systems to mitigate these concerns [70, 71].

7.2. Recommendations

- Improved Security Measures: The strength of ChatGPT's security relies on every single component. If any supplier or vendor is compromised, the entire system could be at risk. It's important to understand that many hackers are actively seeking to exploit this solution, increasing the security risk for businesses using ChatGPT. Therefore, organizations need to establish robust encryption protocols and authentication methods to protect user data and prevent cyber threats.
- Continuous Evaluation and Monitoring: This is for the ethical implications of ChatGPT AI chatbot to detect and minimize any possible risks, this is essential due to ChatGPT's self-learning ability. It's crucial to verify that the system functions within set parameters and isn't exploited for malicious purposes. Given its access to vast amounts of data, there's a potential for security breaches if protective measures are insufficient. This could result in the exposure of sensitive data or its exploitation by malicious actors. Hence, strong access controls, change management, and logging systems are vital.
- The need for transparency: Transparency is vital for establishing trust and preventing privacy concerns in the use of ChatGPT. OpenAI's privacy policy addresses data storage, management, and processing, but the lack of stringent data protection measures and regulations exacerbates privacy worries. OpenAI's opaque operations complicate audits and verification, making it challenging to detect and mitigate privacy risks. This lack of transparency underscores the necessity for stronger data protection measures and regulatory supervision.
- Integration of Ethical Standards: Developers should follow ethical standards when creating AI-powered chat applications, emphasizing that they prioritize user privacy and welfare.
- Continuous Innovation: Developers must keep up with new developments in AI and chat technology to improve ChatGPT's capabilities and offer users a more sophisticated and smooth experience. Additionally, developers, legislators, and users must act as an application of the ChatGPT AI chatbot.

8. Conclusion

In conclusion, ChatGPT, a powerful NLP system, offers several advantages in terms of context recognition and generating relevant responses. It supports multiple languages and diverse tones, allowing for flexible communication. By automating chats, ChatGPT saves time and resources while enabling faster dialogue interactions. Businesses can leverage ChatGPT to efficiently respond to customer queries, providing a personalized experience. The sophisticated AI used in ChatGPT enhances customer service and productivity, enabling companies to focus on core tasks and expand their operations.

However, it is important to address the security vulnerabilities and potential attacks associated with ChatGPT. This research highlights various types of attacks, including unauthorized code execution and insufficient access control. To enhance the security level of ChatGPT, cryptographic and non-cryptographic methods are suggested. While ethical and safety considerations remain crucial in the development of conversational AI systems, there are still flaws and potential attacks that need to be addressed. In summary, while ChatGPT offers transformative capabilities in AI and security, it is essential to address the ethical, legal, and societal implications it presents. By actively addressing these challenges, future work should explore additional types of attacks and propose preventive measures to mitigate their occurrence in ChatGPT. Researchers and developers can ensure the responsible and beneficial deployment of ChatGPT in various applications while safeguarding against potential risks.

9. Declarations

9.1. Author Contributions

Conceptualization, A.A. and N.E.; methodology, A.A , N.E., and Z.A.; validation, A.A., Z.A., N.E., M.I.A., and F.A.; formal analysis, A.A. and N.E.; investigation, N.E., Z.A., F.A., M.A., M.I.A., and A.A.; resources, N.E., D.A., A.A., and Z.A.; writing—original draft preparation, All authors.; writing—review and editing, A.A.; Z.A., M.A., and N.E.; visualization, S.A., D.A., F.A., and M.I.A.; supervision, A.A.; project administration, N.E., A.A., and Z.A.; funding acquisition, S.A. and M.I.A. All authors have read and agreed to the published version of the manuscript.

9.2. Data Availability Statement

Data sharing is not applicable to this article.

9.3. Funding and Acknowledgements

We would like to thank the SAUDI ARAMCO Cybersecurity Chair for funding this paper.

9.4. Institutional Review Board Statement

Not applicable.

9.5. Informed Consent Statement

Not applicable.

9.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

10. References

[1] OpenAI. (2024). OpenAI, California, United States. Available online: https://openai.com/ (accessed on January 2024).

[2] Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., ... & Hu, X. (2023). Harnessing the power of LLMS in practice: A survey on chatgpt and beyond. ACM Transactions on Knowledge Discovery, 1-30. doi:10.1145/3649506.

[3] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 Technical Report. arXiv preprint, arXiv:2303.08774. doi:10.48550/arXiv.2303.08774.

[4] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971. doi:10.48550/arXiv.2302.13971.

[5] Gallifant, J., Fiske, A., Levites Strekalova, Y. A., Osorio-Valencia, J. S., Parke, R., Mwavu, R., ... & Pierce, R. (2024). Peer review of GPT-4 technical report and systems card. PLOS Digital Health, 3(1), e0000417. doi:10.1371/journal.pdig.0000417.

[6] Zhu, K., Wang, J., Zhou, J., Wang, Z., Chen, H., Wang, Y., ... & Xie, X. (2023). PromptBench: Towards Evaluating the Robustness of Large Language Models on Adversarial Prompts. arXiv preprint. doi:10.48550/arXiv.2306.04528.

[7] Spatharioti, S. E., Rothschild, D. M., Goldstein, D. G., & Hofman, J. M. (2023). Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment. arXiv preprint. doi:10.48550/arXiv.2307.03744.

[8] Yao, B., Jiang, M., Yang, D., & Hu, J. (2023). Empowering LLM-based Machine Translation with Cultural Awareness. arXiv preprint, arXiv.2305.14328. doi:10.48550/arXiv.2305.14328

[9] Karpinska, M., & Iyyer, M. (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist. Conference on Machine Translation – Proceedings, 406–438. doi:10.18653/v1/2023.wmt-1.41.

[10] Jain, R., Gervasoni, N., Ndhlovu, M., & Rawat, S. (2023). A Code Centric Evaluation of C/C++ Vulnerability Datasets for Deep Learning Based Vulnerability Detection Techniques. ACM International Conference Proceeding Series, 6, 1-10. doi:10.1145/3578527.3578530.

[11] Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. Nature Medicine, 29(8), 1930–1940. doi:10.1038/s41591-023-02448-8.

[12] Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A Large Language Model for Finance. arXiv preprint. doi:10.48550/arXiv.2303.17564.

[13] Mbakwe, A. B., Lourentzou, I., Celi, L. A., Mechanic, O. J., & Dagan, A. (2023). ChatGPT passing USMLE shines a spotlight on the flaws of medical education. PLOS Digital Health, 2(2), e0000205. doi:10.1371/journal.pdig.0000205.

[14] Abdullah, M., Madain, A., & Jararweh, Y. (2022). ChatGPT: Fundamentals, Applications and Social Impacts. 9th International Conference on Social Networks Analysis, Management and Security, 1–8. doi:10.1109/SNAMS58071.2022.10062688.

[15] Curtis, N. (2023). To ChatGPT or not to ChatGPT? The Impact of Artificial Intelligence on Academic Publishing. Pediatric Infectious Disease Journal, 42(4), 275. doi:10.1097/INF.0000000000003852.

[16] Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Switzerland), 11(6), 887. doi:10.3390/healthcare11060887.

[17] Du, H., Teng, S., Chen, H., Ma, J., Wang, X., Gou, C., Li, B., Ma, S., Miao, Q., Na, X., Ye, P., Zhang, H., Luo, G., & Wang, F. Y. (2023). Chat With ChatGPT on Intelligent Vehicles: An IEEE TIV Perspective. IEEE Transactions on Intelligent Vehicles, 8(3), 2020–2026. doi:10.1109/TIV.2023.3253281.

[18] Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. Sen, Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., … Gabriel, I. (2022). Taxonomy of Risks posed by Language Models. ACM International Conference Proceeding Series, 214–229. doi:10.1145/3531146.3533088.

[19] Pearce, H., Tan, B., Ahmad, B., Karri, R., & Dolan-Gavitt, B. (2023). Examining Zero-Shot Vulnerability Repair with Large Language Models. Proceedings - IEEE Symposium on Security and Privacy, 2339–2356. doi:10.1109/SP46215.2023.10179324.

[20] Wu, X., Duan, R., & Ni, J. (2023). Unveiling security, privacy, and ethical concerns of ChatGPT. Journal of Information and Intelligence, 1-14. doi:10.1016/j.jiixd.2023.10.007.

[21] Qammar, A., Wang, H., Ding, J., Naouri, A., Daneshmand, M., & Ning, H. (2023). Chatbots to ChatGPT in a Cybersecurity Space: Evolution, Vulnerabilities, Attacks, Challenges, and Future Recommendations. arXiv preprint, 1-17. doi:10.48550/arXiv.2306.09255.

[22] Tan, T. F., Thirunavukarasu, A. J., Campbell, J. P., Keane, P. A., Pasquale, L. R., Abramoff, M. D., ... & Ting, D. S. W. (2023). Generative artificial intelligence through ChatGPT and other large language models in ophthalmology: clinical applications and challenges. Ophthalmology Science, 3(4), 100394. doi:10.1016/j.xops.2023.100394.

[23] Khoury, R., Avila, A. R., Brunelle, J., & Camara, B. M. (2024). How Secure is Code Generated by ChatGPT? Honolulu, United States. doi:10.1109/smc53992.2023.10394237.

[24] Renaud, K., Warkentin, M., & Westerman, G. (2023). From ChatGPT to HackGPT: Meeting the Cybersecurity Threat of Generative AI. MIT Sloan Management Review, 64428, 5.

[25] Derner, E., & Batistič, K. (2023). Beyond the Safeguards: Exploring the Security Risks of ChatGPT. arXiv preprint. doi:10.48550/arXiv.2305.08005.

[26] 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. doi:10.4018/IJSPPC.320225.

[27] Sebastian, G. (2023). Privacy and Data Protection in ChatGPT and Other AI Chatbots. International Journal of Security and Privacy in Pervasive Computing, 15(1), 1–14. doi:10.4018/ijsppc.325475.

[28] Esmailzadeh, Y. (2023). Potential Risks of ChatGPT: Implications for Counterterrorism and International Security. International Journal of Multicultural and Multireligious Understanding, 10(4), 535–543. doi:10.18415/ijmmu.v10i4.4590.

[29] Aiyappa, R., An, J., Kwak, H., & Ahn, Y. Y. (2023). Can we trust the evaluation on ChatGPT? Proceedings of the Annual Meeting of the Association for Computational Linguistics, 47–54. doi:10.18653/v1/2023.trustnlp-1.5.

[30] Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. Applied Sciences (Switzerland), 13(9), 5783. doi:10.3390/app13095783.

[31] Li, J., Yang, Y., Wu, Z., Vydiswaran, V. G. V., & Xiao, C. (2023). ChatGPT as an Attack Tool: Stealthy Textual Backdoor Attack via Blackbox Generative Model Trigger. arXiv preprint. doi:10.48550/arXiv.2304.14475

[32] Lande, D., & Strashnoy, L. (2023). Causality Network Formation with ChatGPT. SSRN Electronic Journal, 1-16. doi:10.2139/ssrn.4464477.

[33] Sobania, D., Briesch, M., Hanna, C., & Petke, J. (2023). An Analysis of the Automatic Bug Fixing Performance of ChatGPT. Proceedings - IEEE/ACM International Workshop on Automated Program Repair, 23–30. doi:10.1109/APR59189.2023.00012.

[34] Sarel, R. (2023). Restraining ChatGPT. SSRN Electronic Journal, 1-65. doi:10.2139/ssrn.4354486.

[35] Shahriar, S., Allana, S., Hazratifard, S. M., & Dara, R. (2023). A Survey of Privacy Risks and Mitigation Strategies in the Artificial Intelligence Life Cycle. IEEE Access, 11, 61829–61854. doi:10.1109/ACCESS.2023.3287195.

[36] Addington, S. (2023). ChatGPT: Cyber Security Threats and Countermeasures. SSRN Electronic Journal, 1-12. doi:10.2139/ssrn.4425678.

[37] Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. Computers in Biology and Medicine, 158. doi:10.1016/j.compbiomed.2023.106848.

[38] Hastings, M. C. (2021). Secure Multi-Party Computation in Practice. Doctoral Dissertation, University of Pennsylvania, Pennsylvania, United States.

[39] Rajan, A. A., & Rajan, A. A. (2020). Data Anonymization Techniques for Preserving Privacy in Public Release Data Model A Technical Review. International Journal of Scientific Research in Computer Science and Engineering, 8(1), 58–62. doi:10.26438/ijsrcse/v8i1.5862.

[40] Jiang, B., Li, J., Yue, G., & Song, H. (2021). Differential Privacy for Industrial Internet of Things: Opportunities, Applications, and Challenges. IEEE Internet of Things Journal, 8(13), 10430–10451. doi:10.1109/JIOT.2021.3057419.

[41] Bai, T., Luo, J., Zhao, J., Wen, B., & Wang, Q. (2021). Recent Advances in Adversarial Training for Adversarial Robustness. IJCAI International Joint Conference on Artificial Intelligence, 4312–4321. doi:10.24963/ijcai.2021/591.

[42] Zhao, W., Alwidian, S., & Mahmoud, Q. H. (2022). Adversarial Training Methods for Deep Learning: A Systematic Review. Algorithms, 15(8), 283. doi:10.3390/a15080283.

[43] Malle, B., Schrittwieser, S., Kieseberg, P., & Holzinger, A. (2016). Privacy Aware Machine Learning and the Right to be forgotten. ERCIM News, 107(10), 22–23.

[44] Park, J., & Lim, H. (2022). Privacy-Preserving Federated Learning Using Homomorphic Encryption. Applied Sciences (Switzerland), 12(2), 734. doi:10.3390/app12020734.

[45] Angulo, E., Márquez, J., & Villanueva-Polanco, R. (2023). Training of Classification Models via Federated Learning and Homomorphic Encryption. Sensors, 23(4), 1966. doi:10.3390/s23041966.

[46] Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M. M., Bakhtiari, M., Völker, U., Baumbach, J., ... & Buchholtz, G. (2023). Federated machine learning, privacy-enhancing technologies, and data protection laws in medical research: Scoping review. Journal of Medical Internet Research, 25, e41588. doi:10.2196/41588.

[47] Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, 3, 121–154. doi:10.1016/j.iotcps.2023.04.003.

[48] Alawida, M., Mejri, S., Mehmood, A., Chikhaoui, B., & Isaac Abiodun, O. (2023). A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity. Information (Switzerland), 14(8), 462. doi:10.3390/info14080462.

[49] Sharma, S., Ahmed, S., Naseem, M., Alnumay, W. S., Singh, S., & Cho, G. H. (2021). A survey on applications of artificial intelligence for pre-parametric project cost and soil shear-strength estimation in construction and geotechnical engineering. Sensors (Switzerland), 21(2), 1–44. doi:10.3390/s21020463.

[50] Al-Mushayt, O. S. (2019). Automating E-Government Services with Artificial Intelligence. IEEE Access, 7, 146821–146829. doi:10.1109/ACCESS.2019.2946204.

[51] Benzaïd, C., & Taleb, T. (2020). AI for beyond 5G Networks: A Cyber-Security Defense or Offense Enabler? IEEE Network, 34(6), 140–147. doi:10.1109/MNET.011.2000088.

[52] Ahmed, A., Aziz, S., Abd-Alrazaq, A., Farooq, F., Househ, M., & Sheikh, J. (2023). The Effectiveness of Wearable Devices Using Artificial Intelligence for Blood Glucose Level Forecasting or Prediction: Systematic Review. Journal of Medical Internet Research, 25, 40259. doi:10.2196/40259.

[53] Choraś, M., & Woźniak, M. (2022). The double-edged sword of AI: Ethical Adversarial Attacks to counter artificial intelligence for crime. AI and Ethics, 2(4), 631–634. doi:10.1007/s43681-021-00113-9.

[54] Adadi, A., & Berrada, M. (2020). Explainable AI for Healthcare: From Black Box to Interpretable Models. Advances in Intelligent Systems and Computing, 1076, 327–337. doi:10.1007/978-981-15-0947-6_31.

[55] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 1-11.

[56] Xu, P., Zhu, X., & Clifton, D. A. (2023). Multimodal Learning with Transformers: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(10), 12113–12132. doi:10.1109/TPAMI.2023.3275156.

[57] Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. Neural Networks, 113, 54–71. doi:10.1016/j.neunet.2019.01.012.

[58] Rahman, W., Hasan, M. K., Lee, S., Zadeh, A., Mao, C., Morency, L. P., & Hoque, E. (2020). Integrating multimodal information in large pretrained transformers. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2359–2369. doi:10.18653/v1/2020.acl-main.214.

[59] Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2020). Measuring massive multitask language understanding. arXiv preprint, arXiv:2009.03300. doi:10.48550/arXiv.2009.03300.

[60] Sangwan, R. S., Badr, Y., & Srinivasan, S. M. (2023). Cybersecurity for AI Systems: A Survey. Journal of Cybersecurity and Privacy, 3(2), 166–190. doi:10.3390/jcp3020010.

[61] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), 1-210. doi:10.1561/2200000083.

[62] Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277–304. doi:10.1080/15228053.2023.2233814.

[63] Tawfeeq, T. M., Awqati, A. J., & Jasim, Y. A. (2023). The Ethical Implications of ChatGPT AI Chatbot: A Review. Journal of Modern Computing and Engineering Research, 2023, 49–57.

[64] Wang, P. Q. (2024). Personalizing guest experience with generative AI in the hotel industry: there's more to it than meets a Kiwi's eye. Current Issues in Tourism, 1-18. doi:10.1080/13683500.2023.2300030.

[65] Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT–Exploring the ethical issues of an emerging technology. International Journal of Information Management, 74, 102700. doi:10.1016/j.ijinfomgt.2023.102700.

[66] Parra, J. L., & Chatterjee, S. (2024). Social Media and Artificial Intelligence: Critical Conversations and Where Do We Go from Here? Education Sciences, 14(1), 68. doi:10.3390/educsci14010068.

[67] Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. Nature Machine Intelligence, 1(12), 557–560. doi:10.1038/s42256-019-0109-1.

[68] Escobar-Viera, C. G., Porta, G., Coulter, R. W., Martina, J., Goldbach, J., & Rollman, B. L. (2023). A chatbot-delivered intervention for optimizing social media use and reducing perceived isolation among rural-living LGBTQ+ youth: Development, acceptability, usability, satisfaction, and utility. Internet Interventions, 34, 100668. doi:10.1016/j.invent.2023.100668.

[69] Tsai, W. H. S., & Chuan, C. H. (2023). Humanizing Chatbots for Interactive Marketing. The Palgrave Handbook of Interactive Marketing, Springer International Publishing, 255–273. doi:10.1007/978-3-031-14961-0_12.

[70] Kajtazi, M., Holmberg, N., & Sarker, S. (2023). The changing nature of teaching future IS professionals in the era of generative AI. Journal of Information Technology Case and Application Research, 25(4), 415–422. doi:10.1080/15228053.2023.2267330.

[71] Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. IEEE Access, 11, 80218–80245. doi:10.1109/ACCESS.2023.3300381.