

Potential Risks and Prevention Measures in Application of ChatGPT in Medical Diagnosis

Jiaheng Ji^{1*}

¹ School of Information and Science Technology, Northwest University, Xi'an, China

Email Address

136526007@qq.com (Jiaheng Ji)

*Correspondence: 136526007@qq.com

Received: 8 April 2024; Accepted: 23 April 2024; Published: 24 May 2024

Abstract:

ChatGPT, with its powerful natural language processing capabilities, can be used to help communication between doctors and patients, provide medical assistance and personalized health support in medical domain. However, without proper management, there are potential risks associated with using ChatGPT. It may sometimes provide incorrect information or misdiagnose. Information leakage may happen, posing a threat to users' privacy. It difficult to determine who is ultimately responsible for any medical outcomes in the process of using ChatGPT. This paper focuses on the risk of personal privacy leakage and how to prevent it. To prevent the risks, training and education must be enhanced. Localized differential privacy protection technique can be applied to avoid risk of private information leakage. Data validation and cleaning, model training and validation, real-time feedback and adjustment and experts' review and oversight should be implemented to monitor the performance of ChatGPT in medical diagnosis.

Keywords:

ChatGPT, Application Scenarios, Potential Risks, Risk Prevention Measures

1. Introduction

ChatGPT (Chat Generative Pre-trained Transformer) is an artificial intelligence pre-trained language model based on the transformer architecture for natural language processing, developed by the US company Open AI. Through training on large-scale datasets, ChatGPT can generate coherent and logical natural language, produce texts, images, and videos, and answer questions in text. Since ChatGPT has been put into use, a series of studies have been carried out in the medical field, innovating the way of interaction between doctors and patients and applying it to medical decision-making assistance.

Some scholars have already applied ChatGPT to assist in the diagnosis, treatment, and nursing of clinical diseases, summarizing its strengths and weaknesses. One of the largest non-profit medical institutions in the United States, Brigham Hospital in Massachusetts, recently published a research paper on the application of ChatGPT in

clinical medical decision-making. The study showed that ChatGPT achieved a stunning 71.7% accuracy rate in the entire clinical decision-making process. Through careful evaluation, researchers covered various aspects from proposing diagnoses, recommending diagnostic tests, to final diagnosis and nursing management decisions. It's worth noting that ChatGPT performed well in the final diagnosis task, with an accuracy rate of 76.9%. It is pointed out that ChatGPT achieved an overall accuracy of 71.7% in all 36 clinical test questions, demonstrating significant clinical decision-making ability. Particularly noteworthy is its 60.3% accuracy in preliminary differential diagnosis, while achieving the highest accuracy in final diagnosis. Brigham Hospital used ChatGPT to simulate all aspects of the clinical workflow, including diagnostic testing, final diagnosis, and clinical management, through continuous questioning. The result showed that ChatGPT performed well in diagnostic tasks.

Chee [1] utilized ChatGPT for the etiological analysis of dizziness. The results indicated that ChatGPT provided coherent and logical responses, offering general suggestions including supplementing medical history collection, improving examination methods, and implementing symptomatic treatment measures. However, ChatGPT lacked the ability to handle certain nuances in clinical decision-making. Rao and Pang Kim [2] inputted 36 mature cases from clinical manuals into ChatGPT for assisted diagnosis and differential diagnosis. The results showed that ChatGPT achieved an overall accuracy rate of 71.1%. Howard [3] used ChatGPT to assist in selecting antibacterial drugs for infected patients. The answers summarized ChatGPT's understanding of the scenarios and questions, provided options for antibacterial drugs, and included a disclaimer suggesting consultation with an infectious disease expert at the end. Nevertheless, Adetayo.E. Obassa [4] found that ChatGPT could not identify critical factors in the cases, and its recommendations sometimes changed with repeated inquiries, indicating that ChatGPT's situational awareness, reasoning, and consistency needed further improvement.

2. Application Scenarios of ChatGPT in Medical Diagnosis

ChatGPT can be widely applied in medical field in that it possesses powerful natural language processing capabilities.

2.1. Healthcare Communication

ChatGPT's real-time translation function can significantly enhance the efficiency of communication between doctors and patients, with language barriers between doctors and patients overcome [5]. ChatGPT can assist doctors in better understanding the needs and expectations of patients. Through in-depth conversations with patients, ChatGPT can collect and analyze patient feedback, providing doctors with directions and suggestions for improving the quality of medical services.

2.2. Medical Assistance

ChatGPT automatically generates medical record summaries, greatly simplifying the process of storage of consultation records and enabling healthcare professionals to extract key information, thus improving work efficiency. It provides quick answers to patients' questions. ChatGPT automatically reads the core points in the medical literature, helping researchers quickly grasp the main points of the materials and generate a summary for others to view.

2.3. Personalized Health Support

Through dialogue with users, ChatGPT can understand and analyze individual health statuses, lifestyle habits, and potential health problems, thereby providing health advice and support. It can monitor key indicators such as blood pressure and heart rate, by keeping connected with health monitoring devices, and promptly alert users when abnormalities are detected. This real-time health monitoring and warning function help users better understand their physical condition and address potential health problems. ChatGPT can provide users with recommendations and monitor their physical condition and provide health warnings.

3. Potential Risks of Using ChatGPT in Medical Field

While ChatGPT has shown the strength in assisting medical professionals, there are potential risks associated with using it in the medical diagnosis. Incorrect information or misdiagnose may be provided. Personal data privacy is threatened and it is difficult to determine who is ultimately responsible for medical outcomes.

3.1. Risk of Providing Incorrect Information or Misdiagnosis

ChatGPT may sometimes provide incorrect information or misdiagnose. ChatGPT is trained on large datasets, but these datasets may not cover all possible scenarios or medical conditions. As a result, the model may not have encountered similar cases during training and may provide inaccurate responses. It may struggle to fully understand the context of a medical situation, especially when dealing with complex cases or patients with multiple comorbidities. This can lead to incomplete or misleading advice.

ChatGPT is trained on vast amounts of data, but medical data is complex and diverse. Even large-scale training data may not cover all cases and scenarios. Therefore, ChatGPT may misjudge certain rare or complex diseases. While ChatGPT has made significant progress in natural language processing, it still has limitations in understanding and interpreting medical terminology, pathophysiological mechanisms, and complex clinical situations. This can lead to inaccurate diagnoses for some cases. The application of ChatGPT in medical diagnosis often requires the combination of human input and judgment. If the information input by doctors or users is inaccurate or incomplete, or if the understanding of ChatGPT's output is incorrect, it may lead to misdiagnosis.

3.2. Risk of Personal Privacy Leakage

The use of ChatGPT involves the processing of data of patients. The matching algorithm used by ChatGPT may pose threat to personal data privacy. Due to the involvement of vast amounts of personal data in medical institutions, a third party can target ChatGPT itself or related databases to acquire and steal sensitive information, leading to data security risks. There are certain risks associated with transmitting and storing data on the internet. Not only can hackers attack, but technical vulnerabilities, human errors, and similar situations can also potentially lead to the leakage of patient privacy information. Once data is leaked, it can cause losses to individuals and institutions. There have been reports of ChatGPT leaking sensitive information, such as passwords or personal details.

During medical diagnosis, ChatGPT may be used to process and analyze patient information such as medical history, diagnostic results, and treatment plans. If

patients directly input this information into the ChatGPT system, and the system lacks adequate security protection, this information may be accessed by unauthorized third parties. If the ChatGPT system transmits data over insecure network connections, such as public Wi-Fi, the patient's sensitive information may be intercepted or tampered with during transmission. In a medical diagnosis environment, internal personnel of the ChatGPT system may have access to patient data. If these personnel abuse these privileges and access, modify, or delete patient data without authorization, it may lead to privacy leakage. ChatGPT system may integrate with other medical systems or third-party applications to enable data sharing and interaction. If this integration lacks proper security controls, or if data sharing does not adhere to appropriate privacy protection policies, it may result in the leakage of patient data. It is possible that hackers exploit vulnerabilities to attack the system and steal patient data.

3.3. Risk of Delineation of Medical Responsibility

The autonomous nature of ChatGPT's decision-making processes makes it difficult to determine who is ultimately responsible for any medical outcomes. When it is used to assist in diagnosis and treatment recommendations, it challenges ascertaining the boundaries of responsibility. The potential errors or inaccuracies in ChatGPT's outputs add another layer of complexity to the delineation of medical responsibility. If ChatGPT provides incorrect or misleading information that leads to adverse outcomes to patients, it may be difficult to determine whether the fault lies with the AI system, the medical professionals who relied on its output, or both.

4. Risk Prevention Measures

4.1. Enhancing Training and Education

Collecting more medical data and cases, as well as adopting more advanced data processing techniques, can improve the quality and diversity of ChatGPT's training data, which is helpful for reducing the possibility of misdiagnosis. ChatGPT and its underlying algorithms should be regularly updated and validated to ensure their accuracy and reliability. This includes the incorporation of the latest medical knowledge and evidence-based practices.

Medical professionals must be thoroughly trained in the proper use of ChatGPT. This training should cover not only the technical aspects of the system but also the potential risks and limitations of using ChatGPT in medical diagnosis. Strengthening patients' education and awareness of protecting privacy. Patients' privacy education must be enhanced to raise their awareness and importance of privacy protection.

Technical developers need to be aware of the risks and challenges posed by ChatGPT in the field of medical ethics. For instance, when providing medical diagnosis and treatment advice, ChatGPT must take into account issues such as patient privacy protection and information security, while ensuring that the information it provides is accurate, reliable, and lawful. Technical developers need to be equipped with the knowledge to handle these complexities. Furthermore, technical developers must have a thorough understanding of medical ethics regulations, policies, and norms. Doctors should have a comprehensive understanding of concepts, technologies, and ethical challenges related to ChatGPT. Finally, doctors need to maintain a cautious and prudent attitude in using ChatGPT through continuous learning and improvement, respecting patient rights and privacy.

Applying Data Processing Techniques and Localized Differential Privacy Protection Technique

To prevent the risk of personal privacy leakage, ChatGPT system should utilize robust encryption technologies to safeguard patients' data, ensuring it remains protected from unauthorized access or theft during transmission and storage. Additionally, the system should undergo regular security audits and vulnerability scans to promptly identify and address potential security vulnerabilities.

To deal with the risk of private information leakage, localized differential privacy protection technique is introduced, the principle is to add interference data (noise) to the original query results (numerical or discrete numerical), and then return to the third party research institutions; After adding interference, natural persons can

not be located without affecting statistical analysis, so as to prevent personal privacy data leakage.

Table 1 is a sample of data of the patients. Data sets can provide query services for users without disclosing data values.

Table 1. Sample of medical data where a record of 1 indicates that the patient has AIDS, and a record of 0 indicates that does not.

Name	Diagnostic result
Bob	0
Tom	0
Jane	1
Harry	0
David	1

Suppose the user enters parameter i , Call query function $f(i)=\text{count}(i)$ to get the number of record lines with a diagnostic result of 1, and feedback to the user. When an attacker wants to infer whether David is sick and has known that the David record is in line 5, then he can infer the result by count (5)-count(4). If DP is a function satisfies the Differential Privacy protection algorithm, $DP(i)=f(i)+\text{noise}(\text{noise is randomly distributed noise})$. Assuming that $DP(5)$ may output from the set $\{1, 1.5, 2\}$, then $DP(5)$ will output either value with equal probability, making it impossible for an attacker to predict the result by count(5)-count(4), thus protecting personal privacy.

The $\hat{c}$ query result of the original real data after adding noise through Laplace mechanism is shown: $\hat{c} = c + N$ (c : the true value of query original data , N : noise)

The privacy protection level P is defined : $P = 1 - e^{\frac{-w * c * \epsilon}{\Delta f}}$

$$\epsilon = \frac{-\ln (1-p) * \Delta f}{w * c} \quad (1)$$

(w : the measure of the confidence interval ϵ : privacy protection parameter)

According to the formula, based on the w is fixed, the greater p is, the greater ϵ is larger, the higher the availability of data and the true value of the query output is closer.

Experiments: Set $\epsilon = 0.5, 0.1$ and 0.01 , compare the average error values under different value, the results show in Table 2. Discussion: Get that as the privacy parameter goes up, the error goes down, the availability of the data is higher.

Table 2. Set $\varepsilon=0.5, 0.1$ and 0.01 . Compare the average error values under different values.

The Value of ε	Average Error Value
0.5	0.2
0.1	0.34
0.01	1.13

4.2. Implementing Quality Control Mechanisms

It is suggested to implement robust quality control mechanisms to monitor the performance of ChatGPT in real-world medical settings. This can include data validation and cleaning, model training and validation, real-time feedback and adjustment, and experts' review and oversight.

Model Training and Validation: During the training process, techniques such as cross-validation are employed to assess the model's performance and ensure its robustness on unseen data. Additionally, various evaluation metrics, including accuracy, recall, and F1 score, are utilized to comprehensively assess the model's performance. **Real-time Feedback and Adjustment:** It is essential to establish a real-time feedback mechanism to promptly collect user feedback and model performance data. Based on these feedback data, further adjustments and optimizations can be made to enhance the model's accuracy and reliability. **Experts' Review and Oversight:** The suggestions and diagnostic outcomes generated by ChatGPT in medical applications undergo rigorous review and oversight by medical experts.

Strengthening Legal and Ethical Regulation: Formulating corresponding legal and ethical norms can ensure that the application of ChatGPT in medical diagnosis complies with legal and ethical requirements and protects patients' privacy and data security. **Strictly controlling data access permissions:** The ChatGPT system should establish rigorous access control mechanisms to ensure that only authorized personnel can access patients' data. Regular identity verification and access reviews of authorized personnel should also be conducted to prevent unauthorized individuals from obtaining patient data. **Establishing an emergency response mechanism:** The ChatGPT system should establish an emergency response mechanism to address potential privacy breaches. Upon discovering a privacy breach, the system should immediately take measures such as shutting down relevant systems, notifying patients and relevant authorities, to minimize losses and mitigate risks.

5. Conclusions

ChatGPT may be used for health care communication, medical assistance and personalized health support, for it possesses powerful natural language processing capabilities. However, there are potential risks when used. It may sometimes provide incorrect information or misdiagnose. Information leakage which may happen in the process of using ChatGPT, poses a threat to users' privacy. The autonomous nature of ChatGPT's decision-making processes makes it difficult to determine who is ultimately responsible for any medical outcomes.

In order to prevent these risks, technical developers must take risks, challenges and complexities into consideration. They must have understanding of medical ethics regulations, policies and norms. Doctors should have a comprehensive understanding of concepts, technologies, and ethical challenges related to ChatGPT, maintaining a cautious and prudent attitude in using ChatGPT. Localized differential privacy protection technique can be applied to avoid risk of private information leakage. It is

suggested to implement data validation and cleaning, model training and validation, real-time feedback and adjustment and experts' review and oversight to monitor the performance. ChatGPT should be used as a supportive tool, rather than a replacement for human doctors. Doctors should retain the final decision-making authority, especially in complex or critical cases. If we take appropriate measures to avoid risks, the application of ChatGPT can continue to progress, and its role in medical diagnosis can become more prominent.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

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

References

- [1] Chee, J. et al. "Vertigo, likely peripheral": the dizzying rise of ChatGPT. *Eur Arch Otorhinolaryngol*, 2023, 280(10), 4687-4689.
- [2] Rao, A.; Pang, M.; Kim, J. Assessing the utility of ChatGPT throughout the entire clinical workflow: development and usability study. *Med Internet Res*, 2023, 307(3), e230312.
- [3] Howard, A.; Hope, W.; Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect Dis*, 2023, 23(4), 405-406.
- [4] Adetayo, E.O. Responsible application of artificial intelligence in health care. *South African Journal of Science*, 2023, 16(3), 910-916.
- [5] Li, D.Y.; Liu, Q.M. Ethical Risks and Preventive Paths in Application of ChatGPT in Medical Field. *Chinese Medical Ethics*, 2023, 30(10), 1067-1074.



© 2024 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)