The Optimization Strategy and Clinical Application Effect Evaluation of ChatGPT in Medical Diagnosis Assistance

Bohan Qiu

School of Information Technology, Beijing Institute of Technology Zhuhai, Zhuhai, China

1125089695@qq.com

Abstract: With the rapid development of artificial intelligence technology, large language models, such as ChatGPT, are demonstrating their potential in the medical and health field, offering unique value in assisting medical diagnosis and emerging as a frontier hotspot in smart medical research. This research aims to systematically explore ChatGPT's optimization strategies for medical diagnosis assistance and to comprehensively evaluate its clinical application outcomes. Firstly, this paper analyzes the technical principles of ChatGPT and its current application status, and examines its limitations in practice. To meet these challenges, it proposes optimization strategies, including improving technical standards, upgrading medical professional standards, and strengthening ethical and legal safeguards. Finally, by constructing an evaluation system, an empirical analysis of ChatGPT's clinical effects was conducted from the perspectives of diagnostic accuracy and efficiency. The results show that ChatGPT can significantly improve the diagnostic efficiency and resource utilization under the condition of effective optimization and standardized use. However, its diagnostic accuracy will change with the type of disease, and it needs the supervision of doctors when using it. This study provides theoretical reference and practical guidance for the safe and effective use of ChatGPT in clinical environment.

1. Introduction

1.1 Research Background

Recent advancements in artificial intelligence have transformed the medical field. The large language model ChatGPT, based on natural language processing, has garnered significant attention for its application in assisting medical diagnosis. The long-standing problem of uneven resource allocation in the medical industry, and with its efficient text generation and information integration capabilities, ChatGPT demonstrates remarkable advantages in preliminary diagnosis of common diseases, clinical decision support, and automated processing of medical documents. It can swiftly correlate patient symptoms with possible diseases, and ChatGPT can enhance diagnostic efficiency by 20% to 40%, particularly excelling in cancer screening and the differential diagnosis of complex conditions. However, the current application of ChatGPT in the medical field is still constrained by issues such as delayed data updates and limited in-depth clinical reasoning capabilities. Moreover, factors such as diagnostic errors caused by bias in training data, ethical controversies, and risks of patient privacy exposure also hinder its widespread adoption. Therefore, exploring optimization pathways and evaluating clinical effectiveness have become key research topics for the deep integration of AI technology into medical practice.

1.2 Research Significance

The value of this research is multifaceted and can be discussed comprehensively from both theoretical and practical standpoints. Theoretically, this research contributes to enriching and expanding the theoretical framework for applying artificial intelligence in healthcare. It analyzes the technical mechanisms, applicable scenarios, and inherent limitations of ChatGPT in supporting medical diagnosis, thereby laying a solid theoretical foundation and charting a clear course for future investigation. The proposed optimization pathways provide a theoretical underpinning for ongoing

enhancements to AI-powered medical diagnostic systems [1].

Practically, this research significantly enhances the precision and effectiveness of medical diagnosis. By refining ChatGPT's technical capabilities and integrating advanced medical knowledge, the system can better assist clinicians in diagnostic tasks, thereby reducing the likelihood of misdiagnosis and missed diagnoses and improving patient outcomes. An optimized ChatGPT can shorten the time needed to see a doctor, make our medical resources more reasonable, and reduce the pressure caused by insufficient doctors and hospitals. For patients, a more accurate and faster diagnosis process will make them more satisfied with medical services. Furthermore, this study offers valuable insights for navigating the ethical and legal challenges of deploying ChatGPT in clinical settings, fostering robust and sustainable progress in artificial intelligence in the healthcare sector.

2. The Application Status of ChatGPT in Medical Diagnosis Assistance

2.1 The Technical Principle and Characteristics of ChatGPT

The Transformer architecture forms the technological core of ChatGPT, operating through a two-step process. It begins with unsupervised pre-training on vast amounts of internet text, allowing the model to acquire a deep understanding of human language grammar, syntactic structures, and a broad base of general world knowledge. This process builds a robust foundational language model, which is then refined using Reinforcement Learning from Human Feedback (RLHF). This fine-tuning aligns the model more closely with human objectives, improving its ability to understand complex instructions and produce reliable, high-quality responses [2].

In the realm of diagnostic support, ChatGPT offers several key advantages. First, its sophisticated natural language processing allows it to conduct multi-turn, human-like conversations with clinicians or patients, articulating insights with clarity. Second, it can effectively integrate and extract information, rapidly sifting through its extensive internal knowledge base to retrieve and synthesize relevant data. Third, it exhibits remarkable adaptability, capable of adjusting its tone and focus in response to minimal prompts or examples, underscoring its versatility and potential as a valuable tool in a medical setting.

2.2 Application Scenes

2.2.1 Preliminary Diagnosis of Common Diseases

In specific contexts, ChatGPT acts as an "intelligent triage officer" or a "symptom screening tool." Patients describe their own symptoms to it, such as headaches, fevers, and coughs. Relying on its medical knowledge base, ChatGPT generates potential disease hypotheses and conducts a preliminary ranking based on probability levels or urgency [3]. When presented with a fever accompanied by a rash, ChatGPT prompts users to consider multiple possibilities, such as chickenpox, measles, or drug allergies. It prevents unnecessary visits and effectively guides them to the appropriate department, thereby optimizing the allocation of medical resources. It must be emphasized that the "diagnosis" provided is merely a probabilistic conjecture. The reliability of the conclusion depends heavily on the accuracy and completeness of the user's input, and there is a risk of overlooking serious diseases. Therefore, any preliminary judgment provided by ChatGPT must be reviewed and confirmed by a physician and can never serve as the basis for a final diagnosis or treatment plan.

2.2.2 Assist Doctors to Diagnose Complex Diseases

When dealing with rare diseases, complex conditions, or atypical presentations, ChatGPT can serve as a doctor's "knowledge assistant." Doctors input structured or unstructured text information, such as the patient's complete medical history, laboratory test results, and imaging report summaries, into ChatGPT, requesting assistance with differential diagnosis.

Subsequently, the model, leveraging its powerful information association capabilities, quickly lists multiple clinically matching conditions, including some options that the doctor may not have considered, thereby broadening the diagnostic perspective [4]. For example, when faced with a set of complex symptoms involving multi-system damage, ChatGPT might suggest to the doctor to consider

the possibility of an autoimmune disease or a particular genetic syndrome. ChatGPT can rapidly generate interpretations of the clinical significance of specific tests or compile key points from the latest treatment guidelines for related diseases, providing a valuable reference for doctors in formulating treatment plans and saving time on literature review.

2.2.3 Medical Information Retrieval and Knowledge Popularization

ChatGPT demonstrates significant effectiveness and importance in medical information retrieval and patient education. In contrast to conventional search engines, it produces organized, succinct responses by interpreting the intent behind users' natural-language queries. For healthcare professionals, it aids clinical diagnosis by providing details like contraindications for drug combinations, dosage calculations, and standard operating procedures [5]. For the general public and patients, it serves as an around-the-clock digital health consultant, providing clear explanations of disease causes, therapeutic approaches, and post-treatment care advice in accessible language.

For instance, when posed with the query "What daily considerations should a patient with hypertension observe?" it can methodically offer guidance on several fronts, including dietary management (specifically limiting sodium intake), appropriate physical activity, adherence to medication schedules, and regular medical monitoring. This efficient and dependable informational assistance contributes to elevating public health literacy and serves as a valuable tool for cultivating positive doctor-patient relationships.

3. Analysis of the Limitations of ChatGPT in Medical Diagnosis Assistance

3.1 Limitations of Technology

3.1.1 Data Discrepancy and Delayed Update

The knowledge base of ChatGPT is derived from the fixed dataset used during its training, which leads to significant data bias and update latency issues. The training data is sourced from the internet, where information quality is uneven, featuring conflicting viewpoints and including unverified folk remedies. Consequently, the model inadvertently learns and amplifies these biases, leading to the output of erroneous or one-sided medical information [6]. Medicine, as an ever-evolving discipline, is continuously seeing the emergence of new clinical guidelines, drug developments, and disease discoveries (new pathogens). This establishes a clear temporal limitation on ChatGPT's knowledge base, which cannot access and incorporate the latest advancements in real time. The information lag is fatal in a rapidly changing medical environment, as it can lead to diagnostic recommendations that deviate from best practices, posing a potential threat to patient safety.

3.1.2 An Absence of Profound Semantic Comprehension and Rational Analysis

Statistical pattern matching is the basis for ChatGPT to write fluent and professional-looking words, but it doesn't really understand and think. When dealing with medical tasks that require complex causal chain and in-depth logical analysis, its shortcomings become obvious. For example, although this model can accurately tell the symptoms of "angina pectoris" and "muscular infaction", like an experienced doctor, it is difficult to make a strict differential diagnosis according to the subtle differences of patients' symptoms, the changing law of medical history and the internal relations among various examination results. Its answers are sometimes just "sounds reasonable" statements, rather than conclusions based on pathophysiological reasoning. Although it can list all the diseases that can cause chest pain, it can't effectively combine the specific risk factors of patients, the pain radiation area and the dynamic changes of ECG to accurately infer the most fundamental cause. In summary, the lack of logical reasoning capability limits its application value in complex clinical decision-making.

3.2 Lack of Clinical Experience

While possessing a vast repository of academic literature, ChatGPT critically lacks "clinical experience" in real-world application. This experience encompasses the skills honed through

extended practice and the keen perceptiveness required to identify critical diagnostic cues during patient interviews and examinations. Medicine is both a science and an art, demanding the ability to build patient trust, facilitate effective communication, and exercise intuitive judgment in the face of incomplete data. Such tacit knowledge is inherently difficult to fully digitize and integrate into a model. Therefore, when encountering atypical symptoms, patients' vague descriptions or they deliberately conceal information, ChatGPT is easily biased and may miss the real problem [7]. In contrast, experienced doctors can effectively judge the truth of the case through comprehensive observation and interpersonal interaction skills.

3.3 Patient Privacy and Data Security Risks

The application of ChatGPT in a medical-assisted environment has raised serious concerns about patient privacy and data security. When doctors use the model to obtain assistance, the processing, storage, and use of personal identity information and sensitive health information constitute key ethical issues. Developers use the interaction data between users and the model for subsequent model training. It means that sensitive medical record information may be acquired by third parties or stored on publicly accessible servers without sufficient anonymization, creating potential data leakage risks. Even if companies have formulated privacy policies, their actual implementation and transparency are still concerning. In the medical industry, safeguarding patient privacy is both an ethical obligation and a legal requirement. Any improper handling of data could violate patient rights and result in legal issues.

4. Optimization Strategy of ChatGPT in Medical Diagnosis Assistance

4.1 Technology Optimization

4.1.1 Improve Data Processing and Update Mechanism

Building a high-quality, dynamically updated medical data ecosystem requires overcoming data bias issues. The data sources for model training should shift from general internet text to authoritative, structured medical knowledge bases, the latest clinical guidelines (such as NCCN, UpToDate), peer-reviewed journal literature, and standardized electronic medical record databases. Strict quality control and data cleaning must be implemented at the source to ensure the accuracy and reliability of the knowledge. Furthermore, an efficient continuous learning and incremental update mechanism must be established, cooperating with authoritative medical information providers to establish interfaces for monthly or even weekly knowledge synchronization, or to develop a "retrieval-augmented generation" model architecture. In this architecture, ChatGPT acts as a real-time analysis engine, and its answers do not rely solely on internal fixed knowledge but instead instantly retrieve and integrate the latest and most authoritative medical databases to generate responses, ensuring that users always receive information at the forefront of medicine.

4.1.2 Enhance Semantic Understanding and Reasoning Ability

The core of the improvement lies in enhancing the model's medical logical reasoning capabilities, which requires achieving two key technical breakthroughs. First, it is necessary to achieve deep integration of medical domain knowledge, performing continuous pre-training and fine-tuning on vast amounts of medical text data (such as textbooks and clinical practice guidelines). Furthermore, by utilizing knowledge graph technology, complex relationships between entities such as diseases, symptoms, medications, and examinations (e.g., causality, contraindications) should be explicitly integrated into the model's structure, enabling it to transcend mere lexical-level associations and grasp the deep-seated medical logical network. Second, it is recommended to introduce symbolic reasoning mechanisms and a hybrid architecture that combines the pattern recognition strengths of deep learning with the logical reasoning of rule-based symbolic systems. In this way, when the model wants to generate a differential diagnosis list, it can call an internally stored clinical decision tree or diagnostic rule engine maintained by the expert team to logically verify and sort the preliminary results. It can simulate the doctor's thinking process, make the reasoning more transparent, and meet the medical

standards [8].

4.2 Improve Medical Professionalism

To enable a model to move beyond textbook knowledge and handle the complexities of real-world clinical practice, it must be trained on data sources that incorporate "clinical experience." This process requires the systematic collection, anonymization, and curation of a large volume of electronic health records from real-world settings. Such data should encompass the entire patient care journey—from the patient's presenting complaint, history of present illness, and past medical history, through physical exam findings, lab and imaging results, to final diagnoses, treatment plan selections, and long-term prognoses. By leveraging this complete, time-series data to build the model, it can learn to discern nuanced probabilistic relationships between symptoms and diseases, as well as the differential effectiveness of various treatments across diverse patient populations. Critically, this approach can integrate expert physicians' annotations. It captures the cognitive processes experts use to sift through confounding factors, identify salient clinical signs, and arrive at a final diagnosis. This integration greatly enhances the model's capability to make informed decisions in uncertain situations, resulting in recommendations that better reflect the reasoning of a seasoned clinician.

4.3 Ethics and Legal Protection

The development of trustworthy medical AI systems must prioritize privacy protection and data security. Technically, it can be achieved through innovative approaches like federated learning, which allows models to be trained and iterated upon on local hospital servers, with only encrypted model parameter updates being shared. This approach fundamentally mitigates the risk of information leakage inherent in centralized data storage. Furthermore, stringent anonymization protocols must be enforced at every stage of data processing. In the system's design and architecture, a "privacy by design" principle should be implemented, meaning that complete interaction records for individual patients are not retained by default. All data exchanges must employ end-to-end encryption, supported by robust access control mechanisms and comprehensive audit trails. This dual-track strategy combines technical safeguards with procedural discipline to ensure all access to patient data is traceable, thereby protecting patient privacy rights.

5. Evaluation of the Clinical Application Effect of ChatGPT in Medical Diagnosis Assistance

5.1 Diagnostic Accuracy Assessment

In clinical diagnostics, traditional methods rely on a physician's personal experience, patient history, and standard tests, making results susceptible to individual bias and skills. Conversely, ChatGPT, trained on vast amounts of medical information, can integrate diverse data sources to support diagnosis. For instance, in identifying common respiratory conditions like pneumonia, a top-tier hospital's conventional approach achieved about 82% accuracy, which increased to 91% with ChatGPT's assistance. However, when applied to complex and rare diseases, ChatGPT's performance faltered, with accuracy dropping to 72% due to limited data availability, and it performed slightly worse than traditional methods. To sum up, while ChatGPT enhances diagnostic accuracy for common ailments, expert knowledge remains indispensable for handling complex cases.

5.2 Efficiency Improvement Assessment

5.2.1 Save Diagnosis Time

Within standard clinical procedures, medical staff must meticulously review a patient's medical history and test results, a process that typically takes about 30 minutes per diagnosis. By leveraging the capabilities of ChatGPT, healthcare professionals can rapidly filter key information from electronic health records (EHRs), perform a preliminary analysis of lab data, and obtain diagnostic recommendations. In application tests at primary care facilities, when physicians adopted ChatGPT's suggestions for common internal medicine cases, the average consultation time was significantly reduced, thereby improving overall efficiency. Furthermore, in emergency settings for critical

conditions, such as myocardial infarction, ChatGPT can swiftly align patient symptoms with clinical guidelines to aid clinicians in making initial assessments. This reduction in diagnosis time secures a crucial window for emergency interventions.

5.2.2 Improve the Efficiency of Medical Resource Utilization

In areas where doctors and medical equipment are scarce, utilizing ChatGPT to facilitate virtual consultations can help address the issue of insufficient resources. For example, in a remote township health center, there used to be only 50 patients a day because there were too few doctors. After using ChatGPT, patients with mild symptoms can conduct preliminary examinations and receive advice, allowing doctors to focus on those with more serious and complicated conditions. Now we can see 75 patients every day, and the waiting time in line has become shorter. In the use of examination resources, ChatGPT can accurately determine which necessary examinations need to be conducted based on patient symptoms, thereby avoiding unnecessary examination requirements. The data shows that with its help, the resources of examination equipment are saved, the cost of patients' medical treatment is reduced, and the efficiency of medical resources is improved.

6. Conclusion

Through in-depth analysis, we summarize the following viewpoints. As a large language model, ChatGPT exhibits both application prospects and deficiencies in medical diagnostic support. It requires improvements in multiple aspects, including the aspects such as technology, expertise, and ethical norms.

The in-depth application of ChatGPT in the field of medical diagnosis should follow a rigorous improvement strategy: from a technical perspective, it is necessary to create high-level, dynamically updated medical professional databases and develop more advanced medical reasoning mechanisms. From a professional perspective, we must strengthen cooperation with clinicians and incorporate all types of diagnostic information, treatment protocols, and practical experience into model training. At the legal and moral level, we urgently need to establish strict data protection rules and strengthen the supervision mechanism.

To sum up, ChatGPT is a highly useful auxiliary tool. Due to continued technological progress and strong teamwork in healthcare, ChatGPT is expected to improve precision medicine by making high-quality medical resources more accessible to a larger group of people.

References

[1] Zaichik M, Mahmoudzadeh R, Selvan K, et al. Assessing the Potential of AI-Driven Drug Repurposing in Ophthalmology: An Analysis of ChatGPT's Therapeutic Recommendations[J]. Investigative Ophthalmology & Visual Science, 2025, 66(8):2.

[2] Johnson D, Goodman R, Patrinely J, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model[J/OL]. Research Square, 2023.

[3] Garg R K, Urs V L, Agarwal A A, et al. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review[J]. Health Promotion Perspectives, 2023, 13(3): 183.

[4] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns[C]//Healthcare. MDPI, 2023, 11(6): 887.

[5] Lee H. The rise of ChatGPT: Exploring its potential in medical education[J]. Anatomical sciences education, 2024, 17(5): 926-931.

[6] Caruccio L, Cirillo S, Polese G, et al. Can ChatGPT provide intelligent diagnoses? A comparative study between predictive models and ChatGPT to define a new medical diagnostic bot[J]. Expert Systems with Applications, 2024, 235: 121186.

[7] Javaid M, Haleema A, Singh R P. ChatGPT for healthcare services: An emerging stage for an innovative perspective[J]. BenchCouncil Transactions on Benchmarks, Standards & Evaluations, 2023, 3(1).

[8] Han Z, Battaglia F, Udaiyar A, et al. An explorative assessment of ChatGPT as an aid in medical education: use it with caution[J]. Medical Teacher, 2024, 46(5): 657-664.