Review

Opportunities and Challenges for Large Language Models in Primary Health Care

Journal of Primary Care & Community Health
Volume 16: 1–11

DOI: 10.1177/21501319241312571
journals.sagepub.com/home/jpc
Sage

Hongyang Qin[1,2] and Yuling Tong[1]

Abstract

Primary Health Care (PHC) is the cornerstone of the global health care system and the primary objective for achieving universal health coverage. China's PHC system faces several challenges, including uneven distribution of medical resources, a lack of qualified primary healthcare personnel, an ineffective implementation of the hierarchical medical treatment, and a serious situation regarding the prevention and control of chronic diseases. The rapid advancement of artificial intelligence (AI) technology, large language models (LLMs) demonstrate significant potential in the medical field with their powerful natural language processing and reasoning capabilities, especially in PHC. This review focuses on the various potential applications of LLMs in China's PHC, including health promotion and disease prevention, medical consultation and health management, diagnosis and triage, chronic disease management, and mental health support. Additionally, pragmatic obstacles were analyzed, such as transparency, outcomes misrepresentation, privacy concerns, and social biases. Future development should emphasize interdisciplinary collaboration and resource sharing, ongoing improvements in health equity, and innovative advancements in medical large models. There is a demand to establish a safe, effective, equitable, and flexible ethical and legal framework, along with a robust accountability mechanism, to support the achievement of universal health coverage.

Keywords

primary care, large language models, community health, health promotion, tele-medicine, universal health coverage

Dates received: 15 October 2024; revised: 14 December 2024; accepted: 17 December 2024.

Introduction

Primary health care (PHC) is a comprehensive health care approach that includes health promotion, disease prevention, treatment, rehabilitation, and chronic disease management with the goal of ensuring universal access to high-quality PHC services.[1] Since the release of the "Almaty Declaration" in 1978, PHC has gained global recognition as the central objective for achieving universal health coverage, improved health outcomes and equity through integrated and coordinated medical services.[2]

In China, the PHC system is a fundamental component of the health care framework, and the implementation of "the Healthy China 2030 initiative" and "the 14th Five-Year Plan for National Health" needs to be based on strong PHC. The Chinese health care system is structured into 3 levels: primary care, secondary, and tertiary care. PHC plays an indispensable role in the prevention of basic diseases and community health management. Over the past few decades, China has made remarkable progress in health care services, particularly in the areas of maternal and newborn care, infectious diseases, and primary care access.[3] However, the PHC system continues to face significant challenges. China is currently facing a significant aging population and the urgent and growing public health demands brought about by varying rural healthcare systems, making the situation of chronic disease prevention

[1]The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China
[2]Beigan Street Community Health Service Center, Xiaoshan District, Hangzhou, China

Corresponding Author:
Yuling Tong, Department of General Practice and International Medicine, The Second Affiliated Hospital of Zhejiang University School of Medicine, 88 Jiefang Road, Shangcheng District, Hangzhou City, Zhejiang Province 310009, China.
Email: tongyl0313@zju.edu.cn

and control quite severe.[4] There is also an uneven distribution of medical resources and a lack of qualified primary healthcare personnel, particularly in rural and remote areas, where access to quality medical resources is limited. This unevenness in medical resources and the healthcare system's excessive reliance on hospitals have led to a centralization of disease management, resulting in an ineffective implementation of the hierarchical medical treatment.[5] Furthermore, the continuity and coordination of medical services are insufficient,[6] impeding the effective development of PHC. Additionally, the PHC system still lacks comprehensive coverage in areas such as disease prevention, mental health services, and rehabilitation care.[3] The rise of digital health technologies (DHTs), particularly artificial intelligence (AI), has opened up new opportunities for PHC. As a representative AI technology, large language models (LLMs) utilize natural language processing and generative capacity to improve real-time communication and support data analysis for personalized health management, disease screening, and disease diagnosis.[7,8]

In several major national initiatives, the development and implementation of medical AI have been emphasized to advance the progress of precision medicine and public health applications. Relying on the construction of a biological data system, medical AI represented by LLMs has significant creative implications for the development of PHC. China has elevated "comprehensive prevention and control of chronic diseases" to a national strategy, with technologies like medical large models reshaping the means of chronic disease prevention and control. Continuous new progress is being achieved in the full-process management of "prevention - diagnosis - treatment," emphasizing a prevention-first approach that is centered on people.[4]

The widespread adoption of LLMs in PHC encounters numerous challenges, including a lack of transparency in the model, potential misrepresentation of outcomes, data privacy concerns, and social biases. Moreover, ensuring the safe, reliable, and fair use of LLMs without adding to the workload of primary healthcare personnel is still a challenge that needs thorough investigation and solutions.[9] Consequently, effectively integrating LLMs with the PHC system to enhance diagnostics and treatment efficiency while maintaining quality and fairness of medical services, has become a significant issue in the advancement of medical technology worldwide.

The Technical Principles of Large Language Models

Large language models (LLMs) are *neural networks* of *deep learning* specifically designed to address complex language comprehension and reasoning tasks by processing and generating natural language. The core of LLMs is based on *the Transformer architecture*, that significantly improves the efficiency of handling large-scale textual data. The models showcase their ability to generate grammatically accurate and coherent text through 2 crucial stages: *pre-training* and *fine-tuning.*[10] With the advancement of computational capability and the proliferation of data, LLMs demonstrate exceptional efficacy in large-scale language modeling across a diverse array of natural language processing (NLP) tasks, particularly within the domain of health care.

The representative GPT series demonstrates exceptional proficiency in tasks related to natural language processing with its massive parameter scale. Showing the powerful ability of in-context learning, with tasks and *few-shot* demonstrations specified purely via text interaction with the model,[11] GPT-3 is widely considered to be the first LLM. Notably, GPT-4 has significantly improved its capability to tackle complex tasks and can process multiple types of inputs, including text and images, greatly enhancing its potential applications in the medical field.[12]

In addition, Google's Med-PaLM series is also focused on the medical field, demonstrating excellent capabilities in encoding medical knowledge and answering medical questions.[13] Recently, multimodal large language models have begun to demonstrate their potential impact in health care, with notable examples such as GPT-4, Med-PaLM M, and Gemini series, which can process various types of data, including medical images and genomic data.[12,14,15] This multimodal capability presents innovative solutions for complex tasks in healthcare, allowing LLMs to perform exceptionally well with textual data while also integrating various types of medical data.

Potential Applications of Large Language Models in Primary Health Care

Large language models (LLMs) leverage the capabilities of language processing and reasoning, continuously learning and updating to provide more precise and personalized health services for medical providers and patients in support of health care work. Here are some key applications of large language models in primary health care (PHC).

Health Promotion and Disease Prevention

The essence of PHC lies in promoting health and preventing diseases. The "14th Five-Year Plan for National Health" clearly proposes the implementation of comprehensive strategies for chronic disease prevention and control, emphasizing the core requirement of "health-centered" approaches.[4] By improving health education, personalized health management, and disease screening, LLMs can contribute to the implementation of this national strategy. They can also improve people's health literacy and understanding of diseases, ultimately leading to better health outcomes.[16] Patients receive individualized health guidance through natural language dialogue, covering areas such as diet, exercise, and smoking cessation. This approach allows for earlier access to tailored information based on lifestyle habits and health data through patient empowerment and engagement, as well as timely reminders for patients to take intervention measures in order to prevent the onset and progression of diseases.[17]

Furthermore, LLMs has the potential to assist health care institutions in conducting data analysis to identify and predict high-risk groups. For instance, by analyzing medical history, lifestyle habits, and genetic information, LLMs have the capability to identify high-risk groups with chronic diseases, issue early warnings, and provide guidance for intervention measures. At the National University Health System (NUHS), Singapore, the team has created a tool called CardioSight that can capture and showcase information on cardiovascular risk factors at both individual and geographical levels. When combined with the Chronic Disease Management Program (CHAMP), CardioSight[18] can offer targeted intervention strategies. This personalized approach to disease prevention holds promise for enhancing primary-level public health outcomes and advancing universal health coverage.[19]

Medical Consultation and Health Management

Medical consultation is an important component of PHC. LLMs have the potential to function as a physician's assistant by providing a real-time health advice and initial diagnostic support. Through interactive communication with LLMs, patients can receive symptom-based health advice, thereby addressing the limited availability of PHC resources. Acting as an intermediary between patients and physicians, LLMs streamline health consultation process and provide preliminary diagnostic recommendations[20] to alleviate the shortage of qualified medical personnel in PHC.

LLMs play a crucial role in personalized health management by enabling doctors to create tailored health management plans for patients, monitor their health data in real time, and dynamically adjust management strategies based on this information. Patients can receive telemedicine advice and care guidance, improving their remote care experience and enjoying the convenience brought by telehealth services. Recently, Massachusetts Institute of Technology (MIT) collaborated with Google to develop a Health-LLM[21] that analyzes data collected from wearable sensors, monitors physiological parameters in real time, and reviews health records to make health predictions. This tool reminds patients to take necessary intervention measures, thereby optimizing their health management.

Diagnosis, Triage, and Referral

LLMs have the potential to significantly enhance diagnostic efficiency in PHC by assisting with routine and repetitive medical tasks, such as inputting patient data and summarizing medical information. They also provide clinical decision support by quickly analyzing a large volume of medical documents, electronic health records, and other health data. LLMs can provide preliminary diagnoses based on patients' symptoms and medical history, enhancing the diagnostic efficiency and accuracy of primary healthcare providers. Chat Ella,[22] designed as a tool to assist doctors in diagnosing common chronic diseases, has shown excellent diagnostic capabilities and has received positive feedback regarding its usability. They also offer preliminary diagnoses based on symptoms and medical history, as well as current treatment recommendations supported by relevant medical literature,[23] optimizing treatment plans and enhancing treatment outcomes. By attracting patients with high-quality service, timely care, and cost-effective solutions, it aims to increase the footfall in primary care clinics. Leveraging contracted family doctors services, it provides patients with continuous healthcare services.

By promptly evaluating a patient's condition, LLMs could identify high-risk emergency and critical cases and redirect patients to appropriate specialists or health care facilities, thereby reducing instances of misdiagnosis and inappropriate referrals. Improving the information system of the medical community ensures the immediate exchange and verification of medical data.[5] Multimodal large language models further advance the construction of resource-sharing centers in areas such as medical imaging, medical testing, and electrocardiogram diagnosis. This aids in establishing a two-way referral order for primary examinations, advanced diagnoses, and result recognition,[24] optimizing the referral process,[25] and better implementing the hierarchical medical system to improve the current situation of patients facing difficulties in accessing medical care.

Chronic Disease Management

Chronic disease management is a crucial component of PHC, particularly in China where chronic diseases have become a major health burden.[26] Leveraging long-term patient health data such as electronic health records, family history, and lifestyle information, LLMs can offer personalized chronic disease management plans to patients, along with remote monitoring of key physiological indicators such as blood pressure and blood glucose. This technology enables real-time tracking of patient progression and allows for adjustments to treatment plans based on the latest medical research and individual patient conditions,[27] ultimately leading to a reduction in hospitalizations and emergency interventions. LLMs can enhance the capabilities of chronic disease care in primary healthcare. Data statistics indicate that over 80% of the disease burden in China comes from chronic care needs, and there is a significant shortage of nursing staff.[28] Hippocratic AI[29] has launched a chronic disease virtual nursing assistant supported by LLMs, which reminds patients to take their medications, follow care plans, schedule follow-up appointments, and review medication issues. This meets patients' needs for care, alleviates the burden on medical personnel, and eases the pressure on nursing resources.

In addition, LLMs can predict and prevent complications of chronic diseases by analyzing data. With the help of DeepDR-LLM,[30] a diabetes screening model that integrates LLMs and image-based deep learning, PHC physicians can effectively identify patients with *diabetic retinopathy* who are suitable for referral. Patients are more willing to follow recommendations to visit ophthalmologists and adopt healthier lifestyles. LLMs also hold the potential to provide active rehabilitation guidance and training for patients with functional impairments caused by chronic diseases, thereby improving their quality of life.[31]

Mental Health Support

LLMs also demonstrate significant potential in mental health support. Accelerating the integration of mental health services within primary healthcare and communities is one of the best ways to improve the cost-effectiveness of PHC.[32] Through empathetic interactions, LLMs could provide emotional support to patients and aid them in managing stress, anxiety, and depression associated with their condition. It is capable of monitoring the patient's emotional fluctuations and providing tailored psychological interventions for individuals encountering psychological challenges, including companionship, support, and even therapeutic approaches such as cognitive behavioral therapy or mindfulness training, to facilitate timely alleviation of psychological distress for patients,[33] and provide public mental health education programs to enhance mental health literacy.

Additionally, LLMs could assist mental health professionals in identifying high-risk patients and offering early intervention recommendations. Through monitoring language patterns and emotional changes, LLMs could promptly detect mental health issues in patients, including early signs of schizophrenia or mental disorders,[34] and provide further treatment or referral suggestions. Only a very small number of patients reporting depressive symptoms receive adequate treatment.[35] Soulchat[36] can offer humanized psychological counseling services through multi-turn empathetic dialogues, providing a convenient support avenue for those who cannot access mental health services in a timely manner.

Medical Training and Education

LLMs could serve as a valuable training tool for medical professionals, particularly in the realm of PHC where doctors frequently encounter the need to acquire new knowledge and skills in order to navigate the intricate medical landscape. The key bottleneck in building a strong PHC system is the lack of qualified healthcare personnel.[16] Interactive learning through LLMs is emerging as a viable method to provide doctors with current research findings and clinical guidelines, thereby enhancing their diagnostic and therapeutic capabilities. LLMs potentially generate personalized educational materials tailored to expedite doctors' mastery of specific areas of expertise, such as chronic disease, traditional Chinese medicine, oncology. China's first large language model for traditional Chinese medicine, "Zhongjing,"[37] can generate diagnostic analyses, interactive stories, and explanations of disease causes and mechanisms based on data from traditional Chinese gynecological prescriptions. This can assist doctors in improving their reasoning abilities in traditional Chinese medicine consultations. By simulating case analyses and providing immediate feedback, LLMs also hold promise for augmenting doctors' clinical decision-making abilities and elevating their professional skills. Recently, the Chinese Expert Consensus on Artificial Intelligent General Practitioner(AIGP)[38] has been proposed, which is expected to assist general practitioners in fulfilling the learning requirements of their professional training. Additionally, these models can simulate clinical scenarios, providing practical training and skill development.

Challenges and Limitations

Although large language models (LLMs) have shown significant promise in the primary health care (PHC) field, yet their extensive implementation remains hindered by a multitude of technical, ethical, and legal challenges and limitations.

Transparency and Interpretability

The internal mechanism of LLMs, as deep learning models, is usually considered a "black box," making them challenging to explain the reasoning process in complex medical scenarios. The lack of transparency has a direct impact on the interpretability of models, particularly in the context of medical diagnosis and decision support. It is essential for both doctors and patients to comprehend how the models arrive at specific conclusions. In China, many patients have low trust in primary-level healthcare institutions and weak understanding of the primary care first diagnosis system. The outputs of the model must provide reasonable explanations to enhance trust and guide patients' healthcare beliefs and habits. However, most of the current LLMs, especially commercial models such as GPT-4, have not yet disclosed their specific training data or internal mechanisms.[39]

The research has shown that the use of techniques such as *Chain of Thought (CoT)* can enhance the interpretability of models in PHC, elucidating the reasoning process and facilitating users' comprehension of the logic behind the models' output. Additionally, advancements in more transparent and traceable LLMs contribute to their increased adoption rate in medical settings.[40] However, despite these technological advancements offering some solutions, further research and improvement are necessary to advance the interpretability of LLMs beyond their current exploratory stage.

Hallucinations

LLMs exhibit a proclivity for generating content that is seemingly reasonable but ultimately inaccurate or unverified, commonly referred to as "hallucinations."[41] This inaccurate information may have a substantial impact on patient health in medical environments, especially when utilized without the supervision of professional medical personnel. Depending solely on LLMs diagnostic recommendations may also pose potential risks. Primary-level healthcare institutions bear the important responsibility of the primary care first diagnosis system,[5] which is the core of the hierarchical medical system. This system affects the patients' subsequent initial treatment, triage, and referral, as well as their overall sense of trust. Therefore, it is imperative in PHC to meticulously verify and debug LLMs to ensure the accuracy and reliability of the information they provide.

One approach to mitigate the risk of "hallucinations" is to curate datasets tailored for medical scenarios during the pretraining phase of the model and regularly update these datasets to align with the latest medical research findings.[42] Furthermore, integrating an external knowledge retrieval system to enhance the accuracy of LLMs could effectively reduce distortions and bolster the credibility of LLMs in health care.[43] In resource-scarce western and remote rural areas, developing high-quality LLMs that are suitable for the local primary healthcare system faces significant challenges. These challenges include difficulties in attracting professional talent and collecting and updating training data. To achieve these goals, it is essential to establish effective collaboration and resource sharing with higher-level hospitals, as well as to get government financial incentives and policy support to retain the necessary technical talent. The application of artificial intelligence in healthcare at the grassroots level is continuously being advanced for better and more widespread use.

Data Privacy and Security Concerns

During the training process, LLMs require a significant amount of patient data, including electronic health records (EHR), which contain highly sensitive personal privacy information. Improper use or leakage of this data may result in serious harm to patient rights, particularly as LLMs increasingly support personalized medical services and have the potential to infer specific individual information, thereby posing privacy risks.[44]

To address these issues, the World Health Organization (WHO)[45] supports the enforcement of data protection regulations, requiring that the collection and processing of data must be conducted legally. Additionally, some researchers have proposed the use of privacy-enhancing technologies (PETs) like differential privacy and federated learning while safeguarding personal information. Differential privacy guarantees that the model maintains anonymity when analyzing patient data, while federated learning enables secure sharing and updating of model parameters across multiple distributed devices.[44] Meanwhile, primary-level healthcare institutions must obtain informed consent from patients and adhere strictly to data privacy regulations before utilizing LLMs in order to safeguard the security of patient information.[46]

Bias and Fairness

The training data for LLMs usually reflects societal prejudices based on race, gender, and geographical location, potentially leading to unfairness in medical applications. For instance, certain social groups may lack access to high-quality health advice due to insufficient data or may receive unconsciously biased recommendations from LLMs.[47] This issue is particularly significant in PHC, where the emphasis is on health equity and ensuring equal access to quality medical services for all individuals.[48] China has a diverse cultural background with multiple ethnic groups and is experiencing a significant aging population. It is essential to ensure that minority groups and the elderly have equal access to high-quality healthcare services.

To address these challenges, researchers have proposed *deep reinforcement learning* techniques to mitigate model bias and align it with social ethics.[47] Additionally, it is necessary to introduce cross-cultural datasets and provide intelligent models that cater to dialects and the elderly, ensuring that the models can meet the diverse healthcare needs of patients and avoid injustice stemming from cultural or racial differences.[49] Currently, multimodal large models are placed in "paid access areas,"[45] raising the urgent issue of how to improve the accessibility of these models to cover a broader range of populations.

Human Participation and Ethics Concerns

While LLMs boast powerful automation capabilities, human involvement remains crucial in medical settings. LLMs cannot fully replace doctors, especially in scenarios involving complex ethical decisions or the emotional needs of patients.

Some patients may feel uneasy about relying on artificial intelligence for diagnosis or treatment, particularly in critical moments concerning their health and well-being.[50]

Moreover, it is essential to emphasize the role of LLMs as a supplementary tool rather than a replacement for human decision-making. Researchers have proposed techniques such as *reinforcement learning from human feedback (RLHF)* and *direct preference optimization (DPO)* to align the output of LLMs more closely with human values and preferences, thereby bolstering patient satisfaction and trust.[51] In China, the economic returns for healthcare providers depend on the quantity of medical services provided.[16] As a new type of artificial intelligence technology, how LLMs can encourage healthcare professionals to engage in training and active learning, and how to educate them to correctly handle information from LLMs without relying solely on it, requires further solutions.

Legal and Regulatory Frameworks

The widespread use of LLMs in PHC requires robust legal and regulatory backing. Currently, there is a lack of consistent laws and regulations globally to govern the use of LLMs in health care, which poses potential risks due to the high-stakes nature of medical care. In the context of PHC, the patient harm caused by the use of LLMs cannot be clearly attributed to any party, and there is a lack of regulatory accountability. Therefore, it is essential for LLMs developers, medical institutions, and government regulatory agencies to establish clear lines of responsibility and implement stringent regulatory measures.[52]

Due to issues related to interpretability and safety, some companies' LLMs are unable to comply with existing legal regulations, which restricts their formal operation. This year, the World Health Organization released guidelines on the ethics and governance of large multi-modal models for health to further regulate and standardize the development and application of LLMs, particularly when integrated into the high-risk field of healthcare.[45] The updates and iterations of LLMs are dynamically evolving processes that brings new risks and challenges. Moving forward, global cooperation must be strengthened to ensure that the applications of LLMs align with ethical norms, prioritize patient safety, and promote responsible usage in health care settings.

Future Prospect

The integration of large language models (LLMs) in primary health care (PHC) is anticipated to yield systematic enhancements through optimized resource allocation, improved diagnosis and treatment efficiency, and personalized health management. However, achieving full integration poses several technical, ethical, and legal challenges that require addressing, as well as the promotion of technological innovation and practice. The focus of future prospects can be summarized as follows:

Interdisciplinary Collaboration and Data Integration

The multimodal capabilities of LLMs are poised to play a transformative role in revolutionizing health care applications. In the future, LLMs will increasingly integrate with various digital health technologies, including wearable devices, electronic health records, genomics, to achieve comprehensive health management that covers the entire population and the entire life cycle of individuals. This necessitates close collaboration between medical institutions and technology companies, as well as interdisciplinary cooperation in medicine, data science, and ethics to develop model products for PHC that genuinely meet the requirements of patients and healthcare providers, resulting in the integration of "code to bedside".[53]

The integration of services from primary care to tertiary care, as well as the collaboration between public and private healthcare providers is of crucial significance.[54] The potent data analysis capability of LLMs can facilitate the sharing and standardization of patient health information, and LLMs may turn into the hub of medical data, providing comprehensive clinical care pathways and enabling patients to obtain more convenient 2-way referral services, thereby alleviating the complex care challenges posed by chronic diseases and geriatric diseases.

It is recommended to staff primary-level healthcare institutions with a diverse range of professionals, including medical workers like dentists, traditional Chinese medicine practitioners, and rehabilitation doctors, as well as non-medical personnel such as psychological consultants and community workers. As a mediator for multi-disciplinary cooperation, LLMs can incorporate public services such as elderly care, disability care, and mental health care into primary healthcare services. In rural and remote areas, the addition of multi-disciplinary roles may mitigate the shortage of qualified general practitioners, and the combination of multidisciplinary teams and LLMs can assist in establishing a comprehensive healthcare service centering on primary-level medical services, coordinating the healthcare needs of patients and residents.[55]

Continuous Improvement in Health Equity

The use of LLMs in PHC should prioritize tackling social biases and health disparities. In the future, as more cross-cultural and diverse data are incorporated, LLMs will be better positioned to deliver equitable health services, preventing unfair treatment of specific groups due to constraints of the training data.[56] Furthermore, developers should further enhance the bias correction mechanism of the model to ensure that LLMs offer consistent quality of care to patients from various backgrounds.

Currently, the uneven distribution of medical resources and the imbalanced development between regions and urban-rural areas remain significant challenges hindering the advancement of healthcare in China. Medical artificial intelligence is an effective means to empower primary healthcare.[24] The applications of LLMs are anticipated to significantly enhance telemedicine services, including remote consultations and chronic disease management. This will ensure that primary-level healthcare institutions are accessible across all regions and provide real-time service support, while also fostering remote collaboration with higher-level hospitals. Additionally, it is essential to fully leverage the role of county-level medical communities to promote the descent of high-quality medical resources. By integrating Large Language Model-assisted diagnosis and treatment technologies, the clinical decision-making capabilities of PHC physicians can be enhanced, ultimately improving the quality of medical services.

Furthermore, an effective PHC system requires community outreach.[54] It is crucial to strengthen health promotion efforts beyond medical institutions. Through real-time and convenient health monitoring and education enabled by LLMs, we can enhance health awareness among more residents and encourage them to actively adopt healthy lifestyles. This involves promoting family-based and technology-supported self-health management to proactively prevent diseases and promote health equity.

Construct Ethical and Legal Frameworks

As LLMs continue to be integrated into PHC, enhancing ethical and legal frameworks becomes more vital. At present, major countries and regions around the world are accelerating the improvement of legal supervision related to LLMs to ensure that their development and applications adhere to ethical standards and legal frameworks. For instance, the European Union promulgated the first comprehensive legislation on artificial intelligence, the "Artificial Intelligence Act",[57] which has become the core of the supervision and governance of large models. Combined with other acts in the fields of personal data protection strategies and market supervision, they jointly constitute the supervision system. In the United States, the regulatory framework for large models and the artificial intelligence systems they support is primarily comprised of various legislation at both the federal and state levels. This framework emphasizes the need to balance the rights and interests of enterprises with those of users.[58] The current supervision system for large models in China mainly adopts a strategy of parallel vertical regulations and horizontal supervision tools, reflecting the balance between government leadership and collective interests.[59] In the development of digital medical technologies, including LLMs, there are differences in research transparency, privacy, and medical device approval mechanisms between China and other countries and regions.[60] When the government collects biological data, it may infringe on the privacy of individual patients. However, these privacy data can only be utilized domestically in accordance with compliance regulations. Additionally, there are issues related to the unclear responsibilities of various entities and the absence of regulatory accountability.

In the future, the accountability mechanism will be a significant topic. The government, developing companies, and primary-level healthcare institutions need to clearly define responsibilities to ensure that timely remedial actions can be taken when issues arise during the utilization of LLMs. Collaborative efforts should be made to develop and refine relevant technical and compliance standards. Furthermore, it is advocated to establish flexible and adaptive governance pathways for LLMs to address the new opportunities and challenges brought about by the rapid development of LLMs, adapt to diverse medical scenarios in different socioeconomic backgrounds, continuously monitor and provide feedback in practical applications, and update relevant ethical acts in a timely manner Lastly, issues such as model transparency, privacy, and fairness are universally applicable and require international cooperation and consensus to promote the establishment of a safe, effective, and equitable compliance framework for LLMs.

Continuous Innovation and Long-Term Development

The applications of LLMs in PHC are still evolving, with future innovations aiming to enhance the intelligence of LLMs and expand their range of applications. Ongoing updates and iterations are anticipated to improve the reliability and adaptability of LLMs in handling more complex medical tasks. Develop large models of traditional Chinese medicine with Chinese characteristics, fully leveraging the advantages of traditional Chinese medicine in disease prevention, control, diagnosis, and treatment. Meanwhile, integrate local ethnic medical services in minority regions to better serve the local population. The development of LLMs also requires addressing its flexible application in diverse medical scenarios and promoting it to rural and remote areas. It's critical to identify the health needs of residents in low-resource areas, provide them with information-based infrastructure, and offer policy incentives and economic guarantees to the researchers and medical personnel who develop and use LLMs locally, enabling this new technology to a broader population.

The sustainable development of LLMs necessitates a substantial amount of clinical practice data and long-term follow-up studies. Therefore, further research should prioritize the assessment of the long-term effectiveness of LLMs in PHC, monitoring their impact on patient outcomes, and identifying areas for continuous updates and improvements to ensure their genuine benefit to patients.

Conclusion

Large language models (LLMs) hold great promise as a valuable tool in primary health care (PHC). With their outstanding natural language processing and data analysis capabilities, LLMs can provide support for health care services across various key areas. The effective utilization of LLMs is expected to significantly improve the overall efficiency and quality of PHC, particularly in optimizing resource allocation, enhancing diagnosis and treatment capabilities, and preventing and controlling chronic diseases. To ensure the safe, effective, and fair use of LLMs, it is essential to deepen the technological research and development of medical large models.

Looking ahead, the continued innovation of LLMs in PHC will depend on interdisciplinary collaboration, technology integration, and extensive clinical validation. Through their close integration with digital health technology, LLMs are anticipated to facilitate the modernization of PHC systems and enhance public health, particularly in underserved and remote areas. Accessibility will continue to be a critical factor. Therefore, the global health care community should work together to ensure that LLMs technology can be responsibly applied in medical practice in the future, thereby achieving more efficient and equitable health coverage goals.

Appendix

1. Natural language processing(NLP): NLP is a field of AI research focusing on the interaction between computers and human language. It automates content creation, facilitates language translation, and improves chatbot interactions.
2. Neural networks: Neural networks are computing systems inspired by biological neural networks. It consists of interconnected groups of nodes, often referred to as neurons, which work together to solve specific problems or perform tasks
3. Deep learning: Deep learning is a variant of machine learning involving neural networks with multiple layers of processing nodes, which together facilitate extraction of unstructured input data (eg, images, video, and text).
4. The Transformer architecture: The Transformer architecture is a deep learning model designed for processing sequential data. It utilizes self-attention mechanisms and feedforward neural networks to capture relationships between elements in the input sequence, effectively addressing long-distance dependency issues.
5. Pre-training: Pre-training is the models acquire knowledge of grammar, vocabulary, and language context from a comprehensive corpus. This enables the models to proficiently grasp the fundamental structure and semantics of the language.
6. Fine-tuning: Fine-tuning is the process of taking a pre-trained model and training it on a smaller, task-specific dataset to improve its performance on that particular task.
7. Few-shot : Few-shot is developed to complete tasks with exposure to only a few initial examples of the task, with accurate generalization to unseen examples.
8. Diabetic retinopathy: Diabetic retinopathy is an eye condition that can cause vision loss and blindness in people who have diabetes. It affects blood vessels in the retina (the light-sensitive layer of tissue in the back of your eye).
9. Chain-of-thought(CoT): CoT refers to a reasoning process in which a model or individual arrives at a conclusion through a step-by-step reasoning trajectory when solving complex problems.
10. Deep reinforcement learning: Deep reinforcement learning refers to a powerful framework that allows machines to learn from their environment and improve their performance over time by leveraging the strengths of both reinforcement learning and deep learning.
11. Reinforcement learning from human feedback (RLHF): RLHF uses a reward model to learn alignment from human feedback. This reward model, after being tuned, is able to rate different outputs and score them according to their alignment preferences given by humans. The reward model gives feedback to the original LLM and this feedback is used to tune the LLM further.
12. Direct Preference optimization: DPO is a mapping between reward functions and optimal policies to show that this constrained reward maximization problem can be optimized exactly with a single stage of policy training.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was supported by the National Key Research and Development Program of China(2022YFF0902004), Research and Demonstration Application of Key Technologies for a Homogeneous and Inclusive Healthcare Intelligent Service Platform in Urban and Rural Areas.

ORCID iD

Hongyang Qin https://orcid.org/0009-0004-6577-0070

References

1. World Health Organization and United Nations Children's Fund. A vision for primary health care in the 21st century: towards universal health coverage and the Sustainable Development Goals. 2018. Accessed December 6, 2024. https://www.who.int/publications/i/item/WHO-HIS-SDS-2018.15
2. World Health Organization.Regional Office for Europe. *Declaration of Alma Ata*. World Health Organization. 1978. Accessed December 6, 2024. https://iris.who.int/handle/10665/347879
3. Yip W, Fu H, Jian W, et al. Universal health coverage in China part 1: progress and gaps. *Lancet Public Health*. 2023;8(12):e1025-e1034. doi:10.1016/s2468-2667(23)00254-2
4. Yuan T, Ruan M, Zhu C, Zhang X, Xiong Y, Zhang L. Progress of prevention and control of chronic diseases in 2023. *Life Sci*. 2024;36(1):30-38. doi:10.13376/j.cbls/2024005 (in Chinese)
5. Wu W, Su M. The implementation path of hierarchical medical treatment in China from the perspective of holistic governance. *Inner Mongol Soc Sci*. 2024;45(5):92-99. doi:10.14137/j.cnki.issn1003-5281.2024.05.012 (in Chinese)
6. Li X, Krumholz HM, Yip W, et al. Quality of primary health care in China: challenges and recommendations. *Lancet*. 2020;395(10239):1802-1812. doi:10.1016/s0140-6736(20)30122-7
7. Liu S, Wright AP, Patterson BL, et al. Assessing the value of ChatGPT for clinical decision support optimization. Preprint. Posted online February 23, 2023. medRxiv. doi:10.1101/2023.02.21.23286254
8. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. *J Med Internet Res*. 2021;23(9):e29839. doi:10.2196/29839
9. Thirunavukarasu AJ, Hassan R, Mahmood S, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care. *Jmir Med Educ*. 2023;9:e46599. doi:10.2196/46599
10. Devlin J, Chang M W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. Preprint. Posted online 2018. arXiv. doi:10.48550/arXiv.1810.04805
11. Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. Paper presented at: NIPS'22: 36th International Conference on Neural Information Processing Systems; 28 November-9 December 2022; New Orleans, LA, USA.
12. Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report. Preprint. Posted online 2023. Arxiv. doi:arXiv:2303.08774
13. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. doi:10.1038/s41586-023-06291-2
14. Tu T, Azizi S, Driess D, et al. Towards generalist biomedical AI. *NEJM AI*. 2024;1(3):AIoa2300138. doi:10.1056/AIoa2300138
15. Team G, Anil R, Borgeaud S, et al. Gemini: a family of highly capable multimodal models. Preprint. Posted online 2023. Arxiv. doi:arXiv:2312.11805
16. Yip W. Improving primary healthcare with generative AI. *Nat Med*. 2024;30(10):2727-2728. doi:10.1038/s41591-024-03257-3
17. Brotons C, Björkelund C, Bulc M, et al. Prevention and health promotion in clinical practice: the views of general practitioners. *Prevent Med*. 2005;40(5):595-601. doi:10.1016/j.ypmed.2004.07.020
18. Dalakoti M, Wong S, Lee W, et al. Incorporating AI into cardiovascular diseases prevention–insights from Singapore. *Lancet Reg Health West Pac*. 2024;48:101102. doi:10.1016/j.lanwpc.2024.101102
19. Peckham S, Hann A, Kendall S, Gillam S. Health promotion and disease prevention in general practice and primary care: a scoping study. *Prim Health Care Res Dev*. 2017;18(6):529-540. doi:10.1017/S1463423617000494
20. Song H, Xia Y, Luo Z, et al. Evaluating the performance of different large language models on health consultation and patient education in urolithiasis. *J Med Syst*. 2023;47(1):125. doi:10.1007/s10916-023-02021-3
21. Kim Y, Xu X, McDuff D, Breazeal C, Park HW. Health-llm: large language models for health prediction via wearable sensor data. Paper presented at: Proceedings of the fifth Conference on Health, Inference, and Learning, Cornell Tech, New York City, June 27-28, 2024. doi:arXiv:2401.06866
22. Zhang S, Song J. A chatbot based question and answer system for the auxiliary diagnosis of chronic diseases based on large language model. *Sci Rep*. 2024;14(1):17118. doi:10.1038/s41598-024-67429-4
23. Nashwan AJ, AbuJaber AA. Harnessing the power of large language models (LLMs) for electronic health records (EHRs) optimization. *Cureus*. 2023;15(7):e42634. doi:10.7759/cureus.42634
24. National Health Commission of the People's Republic of China. Transcript of the Press Conference by the National Health Commission on June 18, 2024. Accessed November 15, 2024. http://www.nhc.gov.cn/xcs/s3574/202406/22afc864fc484386b4ab047c7e68f144.shtml (in Chinese)
25. Rose AJ, Timbie JW, Setodji C, Friedberg MW, Malsberger R, Kahn KL. Primary care visit regularity and patient outcomes: an observational study. *J Gen Intern Med*. 2019;34:82-89. doi:10.1007/s11606-018-4718-x
26. Li X, Lu J, Hu S, et al. The primary health-care system in China. *Lancet*. 2017;390(10112):2584-2594. doi:10.1016/S0140-6736(17)33109-4
27. Lázaro E, Yepez J C, Moscardó V, et al. Efficiency of natural language processing as a tool for analysing quality of life in patients with chronic diseases. A systematic review. *Comput Hum Behav Rep*. 2024;14:100407.doi:10.1016/j.chbr.2024.100407

28. Chen W, Ma Y, Yu C. Unmet chronic care needs and insufficient nurse staffing to achieve universal health coverage in China: analysis of the Global Burden of Disease Study 2019. *Int J Nurs Stud*.2023;144:104520. doi:10.1016/j.ijnurstu.2023.104520
29. Webster P. Six ways large language models are changing healthcare. *Nat Med*. 2023;29:2969-2971.doi:10.1038/s41591-023-02700-1
30. Li J, Guan Z, Wang J, et al. Integrated image-based deep learning and language models for primary diabetes care. *Nat Med*. 2024;30:2886-2896. doi:10.1038/s41591-024-03139-8
31. Khalid UB, Naeem M, Stasolla F, Syed MH, Abbas M, Coronato A. Impact of AI-powered solutions in rehabilitation process: recent improvements and future trends. *Int J Gen Med*. 2024;17:943-969. doi:10.2147/IJGM.S453903
32. Xu Z, Gahr M, Xiang Y, Kingdon D, Rüsch N, Wang G. The state of mental health care in China. *Asian J Psychiatry*. 2022;69:102975. doi:10.1016/j.ajp.2021.102975
33. Ma Z, Mei Y, Su Z. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. *AMIA Annu Symp Proc*. 2024;2023:1105-1114.
34. D'Alfonso S. AI in mental health. *Curr Opin Psychol*. 2020;36:112-117. doi:10.1016/j.copsyc.2020.04.005
35. Lu J, Xu X, Huang Y, et al. Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study. *Lancet Psychiatry*. 2021;8(11):981-990. doi:10.1016/S2215-0366(21)00251-0
36. Chen Y, Xing X, Lin J, et al. Soulchat: improving LLMs' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations. Paper present at: Findings of the Association for Computational Linguistics: EMNLP 2023; December 2023; Singapore. Accessed December 6, 2024, pp.1170-1183. Association for Computational Linguistics. https://aclanthology.org/2023.findings-emnlp.83
37. Yang S, Zhao H, Zhu S, et al. Zhongjing: enhancing the Chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. Paper present at: Proceedings of the AAAI Conference on Artificial Intelligence; February 20-27, 2024; Vancouver, Canada. Accessed December 6, 2024. https://ojs.aaai.org/index.php/AAAI/article/view/29907
38. Tsinghua University Vanke School of Public Health, Peking University School of Public Health, Chinese Association of General Practitioners of Chinese Medical Doctor Association, , et al. Chinese expert consensus on artificial intelligence general practitioner. *Chin Gen Pract*. 2025;28(2):135-142. doi:12114/j.issn.1007-9572.2024.0453 (in Chinese)
39. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi M D. Evaluating ChatGPT as an adjunct for radiologic decision-making. Preprint. Posted online February 7, 2023. *MedRxiv*. doi:10.1101/2023.02.02.23285399
40. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. Paper presented at: NIPS'22: 36th International Conference on Neural Information Processing Systems; 28 November-9 December 2022; New Orleans, LA.
41. McKenna N, Li T, Cheng L, Hosseini MJ, Johnson M, Steedman M. Sources of hallucination by large language models on inference tasks. Paper presented at: Findings of the Association for Computational Linguistics: EMNLP 2023, Toronto, Canada, December 6-10, 2023; pp.2758-2774. Association for Computational Linguistics.
42. Zhang Y, Li Y, Cui L, et al. Siren's song in the AI ocean: a survey on hallucination in large language models. Preprint. Posted online September 2023. *arXiv*. doi:arXiv:2309.01219
43. Xia M, Zhang X, Couturier C, Zheng G, Rajmohan S, Rühle V. Hybrid-RACA: hybrid retrieval-augmented composition assistance for real-time text prediction. Paper present at: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track; November 2024; Miami, FL, USA; Accessed December 6, 2024. https://aclanthology.org/2024.emnlp-industry.11
44. Sebastian G. Privacy and data protection in ChatGPT and other AI Chatbots: strategies for securing user information. *Int J Secur Priv Pervasive Comput*. 2023;15(1):1-14. doi:10.4018/IJSPPC.325475
45. World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. 2025. Accessed November 15, 2024. https://www.who.int/publications/i/item/9789240084759
46. Meng X, Yan X, Zhang K, et al. The application of large language models in medicine: a scoping review. *Iscience*. 2024;27(5):109713. doi:10.1016/j.isci.2024.109713
47. Yang J, Soltan AA, Eyre DW, Clifton DA. Algorithmic fairness and bias mitigation for clinical machine learning with deep reinforcement learning. *Nat Mach Intell*. 2023;5(8):884-894. doi:10.1038/s42256-023-00697-3
48. Shi L. The impact of primary care: a focused review. *Scientifica (Cairo)*. 2012;2012:432892. doi:10.6064/2012/432892
49. Cao Y, Zhou L, Lee S, Cabello L, Chen M, Hershcovich D. Assessing cross-cultural alignment between ChatGPT and human societies: an empirical study. Paper presented at: Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP); May 2023; Dubrovnik, Croatia: 53-67. Association for Computational Linguistics.
50. Tyson A, Pasquini G, Spencer A, Funk C. 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Pew Research Center Science & Society. 2023. Accessed November 17, 2024. https://coilink.org/20.500.12592/c3qpj8
51. Minaee S, Mikolov T, Nikzad N, et al. Large language models: a survey. Preprint. 2024. *arXiv*. doi:arXiv:2402.06196
52. Yu P, Xu H, Hu X, Deng C. Leveraging generative AI and large language models: a comprehensive roadmap for healthcare integration. *Healthcare*. 2023;11(20):2776. doi:10.3390/healthcare11202776
53. Yang Z, Silcox C, Sendak M, et al. Advancing primary care with artificial intelligence and machine learning. *Healthc*. 2022;10(1):100594. doi:10.1016/j.hjdsi.2021.100594
54. Tan CC, Lam CS, Matchar DB, Zee YK, Wong JE. Singapore's health-care system: key features, challenges, and shifts. *Lancet*. 2021;398(10305):1091-1104. doi:10.1016/S0140-6736(21)00252-X

55. Ren M, Zhang T, Xu J, et al. Building quality primary health care development in the new era towards universal health coverage: a Beijing initiative. *Global Health Res Policy*. 2024;8(1):53. doi:10.1186/s41256-023-00341-y.
56. Wang X, Sanders HM, Liu Y, et al. ChatGPT: promise and challenges for deployment in low-and middle-income countries. *Lancet Reg Health West Pac*. 2023;41:100905. doi:10.1016/j.lanwpc.2023.100905
57. European Parliament. Resolution on the AI Act. 2023. Accessed November 17, 2024. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf
58. United States Congress. H.R.4223 — 118th Congress (2023-2024). 2023. Accessed November 17, 2024. https://www.congress.gov/bill/118th-congress/house-bill/4223
59. Sheehan M. China's AI regulations and how they get made. *Horizons*. 2023;24:108-125. Accessed November 17, 2024. https://www.jstor.org/stable/48761167
60. Schuerger C, Venkatram V, Quinn K. China and medical AI. Center For Security and Emerging Technology. 2024. Accessed November 17, 2024. https://cset.georgetown.edu/wp-content/uploads/CSET-China-and-Medical-AIImplications-of-Big-Biodata-for-the-Bioeconomy.pdf