*Journal of Evidence-Based Medicine*

WILEY

**REVIEW** OPEN ACCESS

# Large Language Models in Integrative Medicine: Progress, Challenges, and Opportunities

Hiu Fung Yip[1,2] | Zeming Li[3] | Lu Zhang[2,3] | Aiping Lyu[1,2,4]

[1]School of Chinese Medicine, Hong Kong Baptist University, Hong Kong, China | [2]Institute of Systems Medicine and Health Sciences, Hong Kong Baptist University, Hong Kong, China | [3]Department of Computer Science, Hong Kong Baptist University, Hong Kong, China | [4]Guandong-Hong Kong-Macau Joint Lab on Chinese Medicine and Immune Disease Research, Guangzhou, China

**Correspondence:** Lu Zhang (ericluzhang@hkbu.edu.hk) | Aiping Lyu (aipinglu@hkbu.edu.hk)

**Received:** 3 January 2025 | **Revised:** 11 April 2025 | **Accepted:** 5 May 2025

**Keywords:** artificial intelligence | generative artificial intelligence | integrative medicine | Large Language Model | precision medicine

## ABSTRACT

Integrating Traditional Chinese Medicine (TCM) and Modern Medicine faces significant barriers, including the absence of unified frameworks and standardized diagnostic criteria. While Large Language Models (LLMs) in Medicine hold transformative potential to bridge these gaps, their application in integrative medicine remains underexplored and methodologically fragmented. This review systematically examines LLMs' development, deployment, and challenges in harmonizing Modern and TCM practices while identifying actionable strategies to advance this emerging field. This review aimed to provide insight into the following aspects. First, it summarized the existing LLMs in the General Domain, Modern Medicine, and TCM from the perspective of their model structures, number of parameters and domain-specific training data. We highlighted the limitations of existing LLMs in integrative medicine tasks through benchmark experiments and the unique applications of LLMs in Integrative Medicine. We discussed the challenges during the development and proposed possible solutions to mitigate them. This review synthesizes technical insights with practical clinical considerations, providing a roadmap for leveraging LLMs to bridge TCM's empirical wisdom with modern medical systems. These AI-driven synergies could redefine personalized care, optimize therapeutic outcomes, and establish new standards for holistic healthcare innovation.

## 1 | Introduction

Integrative Medicine, which synergizes Traditional Chinese Medicine (TCM) and Modern Medicine, represents a patient-centered paradigm aimed at holistic health assessment and personalized treatment [1, 2]. Rooted in merging disease-pathway analysis with TCM syndrome differentiation, Integrative Medicine has demonstrated clinical efficacy across diverse conditions. Notable examples include artemisinin, a TCM-derived compound validated for tracheitis treatment [3]; acupuncture has been recognized globally for pain management and neurological rehabilitation [4]. Despite these successes, Integrative Medicine remains underutilized due to some systemic barriers. According to the report from the National Health Commission in China in 2022, only 14.6% of China's out of the 8.42 billion medical services involved TCM, with a mere 10–12% delivered by Integrative Medicine practitioners. Even in Hong Kong, 25.9% of patients concurrently use TCM and Modern Medicine [5].

Integrative medicine systems face several obstacles in clinical practice, notably the extensive training required for dual expertise in TCM and Modern Medicine. Those obstacles lead to a

Hiu Fung Yip and Zeming Li contributed equally.

critical shortage of clinicians capable of providing integrated care. Also, unlike a single-discipline medical system, the Integrative Medicine System must address the fusion of decision-making processes from two different medical understandings of diseases. These practical barriers significantly impede the implementation of Integrative Medicine Systems in clinical settings.

Despite these challenges, new opportunities are emerging to overcome the challenge in the Integrative Medicine System. Advanced LLMs like ChatGPT [6] offer the transformative potential to address these gaps. Early applications demonstrate their ability to synthesize cross-paradigm clinical knowledge, generate expert-level insights in glioblastoma management [7], and standardize diagnostic frameworks.

This review systematically evaluates the opportunities and challenges for establishing LLMs in Integrative Medicine, emphasizing the need to address the fusion of decision-making processes from the distinct medical understandings of TCM and Modern Medicine Systems [8, 9]. It highlights the unique applications in integrative medicine, such as the dual translation of clinical reports between TCM and Modern Medicine risk alerts for acute conditions for TCM practitioners and suggestions for TCM alternative treatment plans for doctors [10]. Through experimental evidence, the review demonstrates the limitations of state-of-the-art LLMs in medicine, which are not yet equipped to address tasks unique to integrative medicine. Our work identifies the challenges in developing integrative medicine models, including how to merge the decision-making processes of TCM and Modern Medicine, the scarcity of integrative medicine data, and the lower tolerance for error compared to single-disciplined medical models. Finally, the review proposed a pathway to create LLMs in Integrative Medicine by establishing a common language between TCM and Modern Medicine, leveraging existing evidence-based medicine, molecular biology, and systems medicine research. This work aims to analyze progress, challenges and propose solutions for LLMs in integrative medicine.

## 2 | LLMS and Their Current Progress in Medicine

### 2.1 | LLMs in General Domain

The age of generative AI has come since ChatGPT [11], the third generation of LLMs in General Domain developed by OpenAI, was deployed in 2023. ChatGPT's most impressive capability is its exceptional performance in answering questions, which stems from its utilization of transformer architecture and pretraining that encodes a large corpus into the model. This foundation enables it to provide a chat-like interface for seamless interactions [12]. After that, similar models such as LLaMA [13–15], Qwen [16], Ziya [17], ChatGLM [18], Bloom [19], and Baichuan [20] were proposed by different companies using similar decoder-only transformers (Figure 1). The differences include whether they are open or closed source, the languages available for the user interface, the number of parameters, and the techniques employed to enhance response accuracy. For example, ChatGPT is a closed-source model, whereas the LLaMA series is open-source. Qwen and Alpaca-Chinese are fine-tuned based on LLaMA with adaptability in Chinese, and Bloom is trained to understand more than 13 natural languages. The number of parameters is the key to LLMs' behaviour and performance—usually, the larger the parameter size, the higher the response accuracy. However, increasing the parameters of LLMs also requires scaling up the computational resources needed for training and deploying the model. Hence, the number of parameters for existing LLMs in General Domain commonly ranged from 0.5 to 70 billion after balancing the cost and effectiveness. Regarding the techniques employed to enhance response accuracy, ChatGPT used Reinforcement Learning from Human Feedback (RLHF) to align human feedback with the LLMs response [6]; LLaMA used direct preference optimization (DPO) [21] to improve response accuracy (Figure 2).

LLMs' user-friendly interface and satisfactory response accuracy arouse people's interest in applying them to different fields, including the medical domain. However, pioneering studies have reported LLMs' dissatisfaction with medical queries [22, 23], highlighting the need for further refinement of LLMs to respond to Modern medical queries.

### 2.2 | The Progress and Challenges of LLMs in Modern Medicine

LLMs in Modern Medicine addressed the challenge of suboptimal performance by fine-tuning LLMs in General Domain with specialized medical knowledge. Med-PaLM was proposed by Google's team, which contributes a dataset, HealthSearchQA, for fine-tuning, achieving state-of-the-art performance in MedQA, MedMCQA, PubMedQA, and MMLU and introducing instruction prompt tuning for cost-effective alignment of LLMs in General Domain to medicine knowledge [24]. Many works have been proposed for further advancement (Table 1), but some key limitations remain unsolved.

First, existing LLMs in Modern Medicine fail to capture the dynamic of updated medical knowledge, while factual knowledge in medicine is subject to change over time. The research found that the doubling time of medical knowledge in 1950 was 50 years, and in 2010, 3.5 years, and it is projected to be just 73 days in 2020 [70] (The solution will be discussed in Section 5.1.1).

Second, hallucinations were incredibly toxic in the LLMs in Modern Medicine. Since it would be a disaster for patients to make their next step based on the false opinion of the AI, for domain experts, it would be less toxic since they have sufficient background to evaluate the truthfulness of the answer. However, hallucinations may affect the LLMs that target non-domain experts or junior clinicians (The solution will be discussed in Section 5.1.2).

Third, many of these LLMs in Modern Medicine were trained on English Corpus, which does not cater to practitioners who use Chinese as their primary language. Many works have been proposed to solve the language incompatibility in English trained LLMs, the pioneering work for LLMs. The pioneering work for LLMs in Modern Medicine trained on Chinese Corpus is Zhongjing-LLaMA [39], which argued that the English corpus-trained LLMs might not fully fit into the Modern Medicine context written in Chinese. Hence, Zhongjing-LLaMA teams developed the first LLMs in Modern Medicine trained on Chinese

**TABLE 1** | Summary of existing LLMs in TCM and MM.

| Model | General/TCM/ Modern Medicine | Architecture | Optimization | Params | Version | Data source | Languages | Cite |
|---|---|---|---|---|---|---|---|---|
| Ziya | General | LLaMA | PT & SFT & RLHF | 7B/13B | v1,v1.1 | 110B tokens | zh & en | [25] |
| Qwen | General | LLaMA | PT & SFT & RLHF | 0.5B/1.5B/1.8B/3B/7B/14B/72B | v1/v2.5 | 3T/18T tokens | zh & en & 27 languages | [26] |
| LLaMA | General | Transformer [27] | PT & SFT & RLHF | 7B/8B/13B/33B/34B/65B/70B | v1/v2/v3 | 1.4T/2T/15T tokens | en | [13–15] |
| ChatGLM | General | GLM [28] | PT & SFT & RLHF | 6B | v1/v2 | 1T/1.4T tokens | zh & en | [18] |
| Alpaca-Chinese | General | LLaMA | PT | 7B/13B/33B | v1/v2 | 120G texts/4.3 M instructions | zh | [29] |
| Bloom | General | Megatron-LM GPT2 [30] | PT & SFT | 176B | v1 | 366B tokens | zh & en & 44 languages | [19] |
| Baichuan | General | Transformer | PT & SFT & RLHF & PPO | 7B/13B | v1/v2 | 26T tokens | zh & en | [20] |
| ChatGPT | General | Transformer | PT & SFT & RLHF | Not disclosed | v1/v2/v3/v4 | Not disclosed | zh & en | [6] |
| VisualGLM | General | ChatGLM&BLIP2-Qformer [31] | PT & SFT | 6B | v1 | 30M tokens | zh & en | [28],[32] |
| VisualGPT | General | GPT and encoder-decoder visual learner | PT | Not disclosed | v1 | MS COCO [33] Conceptual Captions [34] and IU X-ray [35] | en | [36] |
| ChiMed-GPT | Modern Medicine | Ziya | PT & SFT & RLHF | 13B | v1 | 369,800 documents medical encyclopedia data<br>8475 articles from medical textbooks | zh & en | [37] |
| Taiyi | Modern Medicine&TCM | Qwen | SFT | 7B | v1 | 170 open-source datasets 38 Chinese datasets covering 10 BioNLP tasks and 102 English datasets covering 12 BioNLP tasks | zh & en | [38] |
| Zhongjing | Modern Medicine | Ziya | PT & SFT & RLHF & PPO | 7B/13B | v1 | CMtMeQA 70,000 multi-round dialogue | zh | [39] |
| BianQue | Modern Medicine | ChatGLM | SFT | 6B | v1/v2 | BianQueCorpus | zh | [40] |
| HuatuoGPT | Modern Medicine | Ziya/Baichuan | SFT | 7B/13B | v1 | HuatuoGPT-sft-data-v1 | zh | [41] |

(Continues)

**TABLE 1** | (Continued)

| Model | General/TCM/ Modern Medicine | Architecture | Optimization | Params | Version | Data source | Languages | Cite |
|---|---|---|---|---|---|---|---|---|
| ShenNong-TCM | TCM | ChatGLM | SFT | 6B | v1 | ShenNong_TCM_Dataset | zh | [42] |
| CareGPT | Modern Medicine&TCM | 26 open sources LLMs | PT & SFT & RLHF | 7B/13B/20B | v1 | 4 PT, 29 SFT, 5 RLHF and 1 distilled dataset | zh | [43] |
| DISC-MedLLM | Modern Medicine | Baichuan | PT & SFT & RLHF & LLM in loop & human in loop | 13B | v1 | DISC-Med-SFT | zh | [44] |
| IvyGPT | Modern Medicine | LLaMA | PT & SFT & RLHF | 33B | v1 | 300K Q&A from subset HuatuoGPT-sft-data-v1 and real doctor-patient conversations | zh | [45] |
| TCMLLM | TCM | ChatGLM | SFT | 6B | v1 | ISGP, CHP, CMCC, and real clinic data | zh | [46] |
| Lingdan | TCM | Baichuan | PT & CoT | 13B | v1 | TCM ancient books, Chinese Pharmacopoeia, Package Insert, Standard Preparation, TCM Case Records, TCM Educational Materials | zh | [47] |
| Med-Gemini | Modern Medicine | Gemini | PT & Web searching& CR [48] | Not disclosed | v1 | Not disclosed | en | [49] |
| PMC-LLaMA | Modern Medicine | LLaMA | PT & SFT | 7B/13B | v1 | MedC-K, MedC-I | en | [50] |
| CPLLM | Modern Medicine | LLaMA, BioMedLM [51] | SFT | 2.7B/13B | v1 | eICU-CRD database [52], Scientific data [14] and MIMIC-IV database [53] | en | [54] |
| Med42 | Modern Medicine | LLaMA | SFT | 7B/70B | v1/v2 | Instruction-tuning dataset | en | [55] |
| MedPaLM-2 | Modern Medicine | GPT-3 [56] | Standard few-shot [57], chain-of-thought [58], self-consistency prompting [59] | Not disclosed | v1/v2 | Medical exam questions [60], MedMCQA [61] | en | [24] |

(Continues)

**TABLE 1** | (Continued)

| Model | General/TCM/ Modern Medicine | Architecture | Optimization | Params | Version | Data source | Languages | Cite |
|---|---|---|---|---|---|---|---|---|
| TCM-GPT | TCM | / | PT & SFT & RAG | Not disclosed | v1 | TCM-Corpus, TCM-Exam, TCM-HER | zh | [62] |
| OYEN | TCM | LLaMA | RAG | Not disclosed | v1 | TCM datasets from Kaggle, UCI machine Learning repository and Hugging Face | zh & en | [63] |
| Qibo | | Chinese-LLaMA | PT & SFT | 7B/13B | v1 | TCM textbooks, TCM ancient Books, encyclopedia of TCM | zh & en | [64] |
| BianCang | Modern Medicine &TCM | Qwen | PT & SFT | 7B | v1 | Textbook, Q&A, Exam, real medical data, and differential diagnoses in both TCM and Modern Medicine based on real hospital records | zh & en | [65] |
| ChatDoctor | Modern Medicine | LLaMA | SFT | 7B | v1 | 52K alpaca and HealthCareMagic100k | en | [66] |
| medAlpaca | Modern Medicine | LLaMA | SFT | 7B/13B/33B/65B | v1 | Medical Meadow including 160K entity | en | [67] |
| DoctorGLM | Modern Medicine | ChatGLM | SFT | 6B | v1 | CMD, MedDialog, HealthcareMagic, and ChatDoctor | zh & en | [68] |
| MedicalGPT | Modern Medicine | Baichuan/ChatGLM/LLaMA/Qwen | PT & SFT & DPO & RLHF & ORPO | 0.5B to 110B | v1 | 2 medical pretrain, 2 general pretrain, 13 SFT, and 9 preference datasets | zh & en | [69] |

Abbreviations: CoT: chain-of-thought, CR: chain-of-reasoning, DPO: direct preference optimization, en: English, ORPO: odd ratio preference optimization, PPO: proximal policy optimization, PT: pretrain, RAG: retrieval augmented generation, RLHF: reinforcement learning with human feedback, SFT: supervised fine-tune, zh: Chinese.

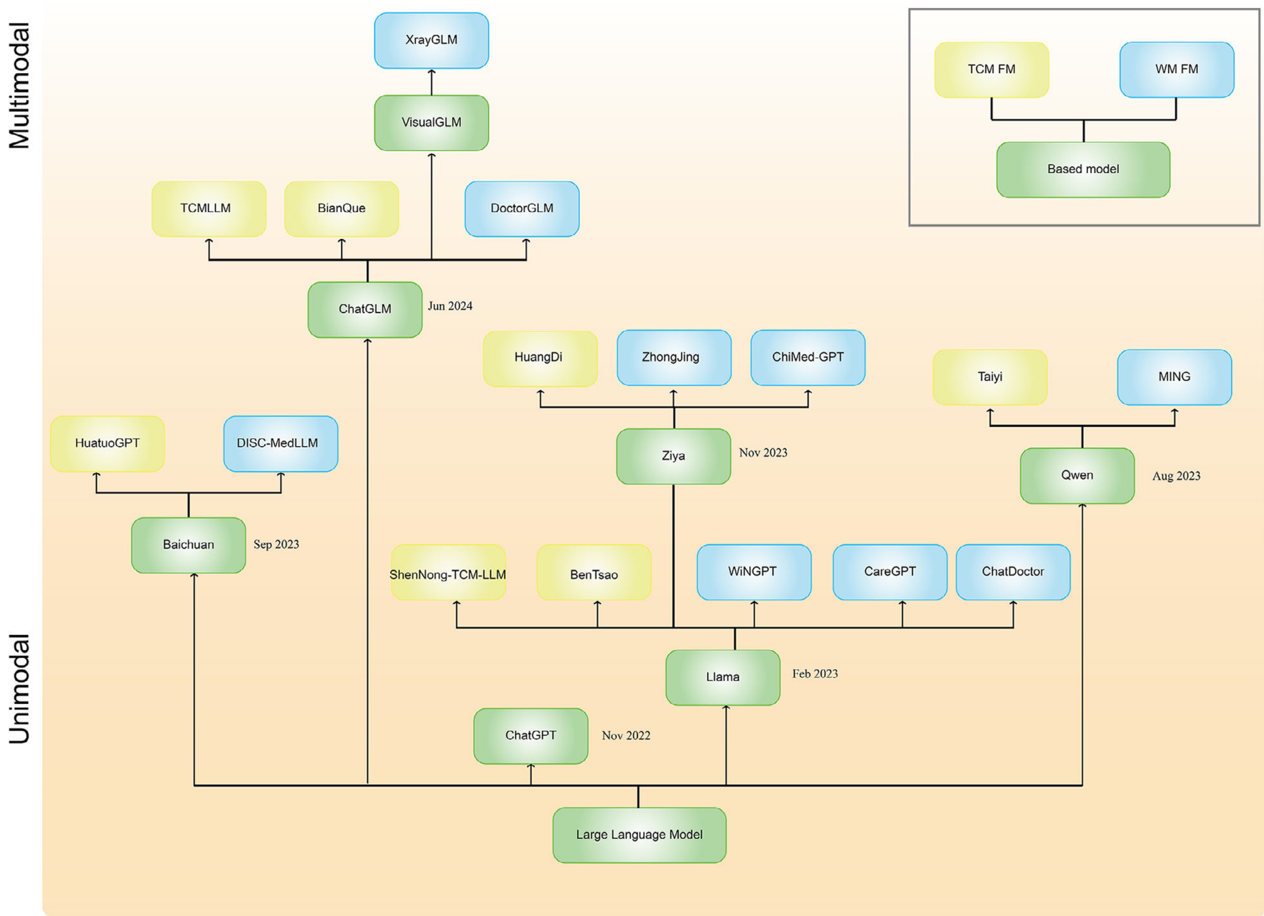


**FIGURE 1** | **Development of Large Language Models in Medicine**.
This figure illustrates the development of LLMs in medicine while green represents a base model without medical domain knowledge, yellow represents fine-tuned with Traditional Chinese Medicine corpus and blue represent fine-tuned with Modern Medicine corpus.

Corpus. They implemented a complete training process involving pretraining, supervised fine-tuning, RLHF, demonstrating excellent generalization capabilities and even approaching the professional level of doctors in specific dialogue scenarios. Later, Tian et al. noticed the need for long text input in medical LLMs, so they utilized the long text adopted structure, Ziya, to build a medical-specific model ChiMed-GPT [37]. Xiong et al. identified the gap in deploying the computational resource-intensive model in clinical practice and proposed a light resource-required model, DoctorGLM [68]. Wang et al. proposed IvyGPT [45] to extend the single round Question and Answering to multi-rounds in order to mimic the situation in a real clinic. Another breakthrough have been made in CareGPT [43], it expanded the model's ability to consider real-world diagnostic processes involving the patient's medical history and ongoing learning to tailor the diagnosis to the individual's specific needs. BianQue [40] identified the limitations in single-turn question-response interactions and proposed "Chain of Questioning" to allow the model to gather more comprehensive data through iterative questioning, enabling it to generate more precise and informative responses.

Since language is not the only approach doctors use to make clinical decisions, models combine multiple modalities to mimic the clinical decision process. For example, XrayGPT [71] incorporated the X-ray and language modalities to simulate real-world clinical decision-making better. HistoGPT was trained to generate high quality report from histopathology images [72]. Moreover, a recent work on nature medicine had shown that LLMs could write spurious clinical text summary compared to physicians [73].

Despite numerous efforts to adapt LLMs for the Chinese context in Modern Medicine, a substantial part of the Chinese medical system, especially TCM, continues to be underrepresented in these models.

### 2.3 | The Progress and Challenges of LLMs in TCM

To address the underrepresentation of TCM in LLMs focused on Modern Medicine, researchers have proposed fine-tuning LLMs in TCM using knowledge from TCM textbooks and herbal formulas. The main difference between LLMs in Modern medicine is that TCM corpus is usually written in Chinese, so the base model (LLMs in General Domain/LLMs in Modern Medicine) to train LLMs in TCM needs to have an introductory Chinese language understanding ability. ShenNong-TCM-LLM pioneered

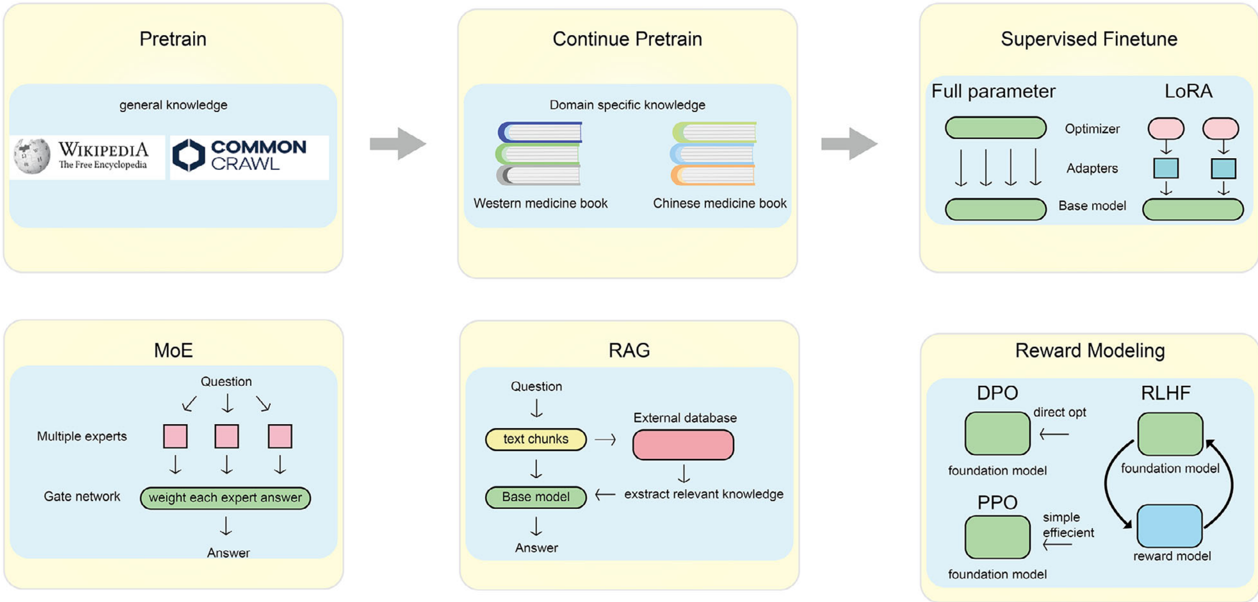


**FIGURE 2** | **Training pipeline of existing Large Language Model**.
This figure illustrates the techniques for training a Large Language Model in medicine including pretraining for incorporate general knowledge, continue pretraining to insert domain specific knowledge, supervised fine-tune is for learning to answer the medical question and answer, reward modeling is for alignment the model response to human experts' response. RAG is conducted to provide rationale and improve accuracy of the response. MoE to emulate scenarios with multiple experts, enhancing response accuracy and specificity.

in this field, which optimized LLaMA with LoRA fine-tuning [42] (Figure 1). It utilized self-instruction to create a tuning dataset comprising eleven thousand TCM-related records. They also incorporated the TCM knowledge graph from their earlier research work into the ShenNong model to improve its response correctness. Later, TCM-LLM [46] integrated real-world clinical records, medical classics, and textbooks of traditional Chinese medicine. Taiyi' [38] extent the monolingual medical question-answering dialogues in either Chinese or English to a bilingual (Chinese and English) model with both TCM and Modern Medicine corpus. More recently, Tianhe Lingshu Large Model specifically trained on data from acupuncture and provided evidence from Chinese medicine texts alongside the response. Lingdan [47] has been released, which can provide answers based on clinical knowledge and offer recommendations for herbal prescriptions, bringing it one step closer to mirroring the complexities of real-world clinical scenarios (Figure 1). Many works have been proposed for further advancement (Table 1), but some key limitations remain unsolved.

First, language barriers due to the complexity and cultural specificity of TCM terminology limit accessibility for practitioners unfamiliar with the concepts. (The solution will be discussed in Section 5.2.)

Second, existing TCM LLMs struggle with engaging in multi-round conversations, which is essential for maintaining coherent and contextually relevant dialogues in clinical settings. (This could be achieved using a similar pipeline for multi-round conversations as LLMs in Modern Medicine.)

Third, the integration of multimodal capabilities—such as visual or auditory data processing—has not yet been realized in TCM models, restricting their ability to conduct comprehensive assessments incorporating various information forms. The solution will be discussed in Section 5.1.4.

Despite numerous efforts to develop LLMs in TCM, its advancement is far behind that in Modern Medicine.

## 2.4 | Uniqueness of LLMs in Integrative Medicine

LLMs in integrative medicine are not simply an ensemble model that merges aspects of Modern Medicine and TCM (Figure 3). The reviewed LLMs in Modern Medicine and TCM provided valuable training pipelines that served as the basis for developing LLMs in Integrative Medicine, such as the long text adopted structure, multi-round conversations, multimodality integration, and Chinese context understanding. It is equally important to anticipate the limitations present in these existing models, such as the dynamic of updated medical knowledge, hallucinations, language barriers from TCM contexts, multimodality and the lack of SOP in clinical practice. We propose solutions to avoid the drawbacks of existing LLMs accumulating in the LLMs in Integrative Medicine (Sections 5.1.1–5.1.4, 5.2.1, 5.3.1, and 5.3.2).

Apart from the general challenges that need to be addressed, the unique characteristics of Integrative Medicine raise some specific challenges and opportunities to improve the existing healthcare system (Table 2). We recognize the opportunities (Section 4) and challenges (Section 5) for LLMs in Integrative Medicine. After identifying them, we propose solutions to address the unique pitfalls in LLMs for integrative medicine (Sections 5.1.4, 5.2.1, and 5.2.2). This approach will ensure a more robust and practical

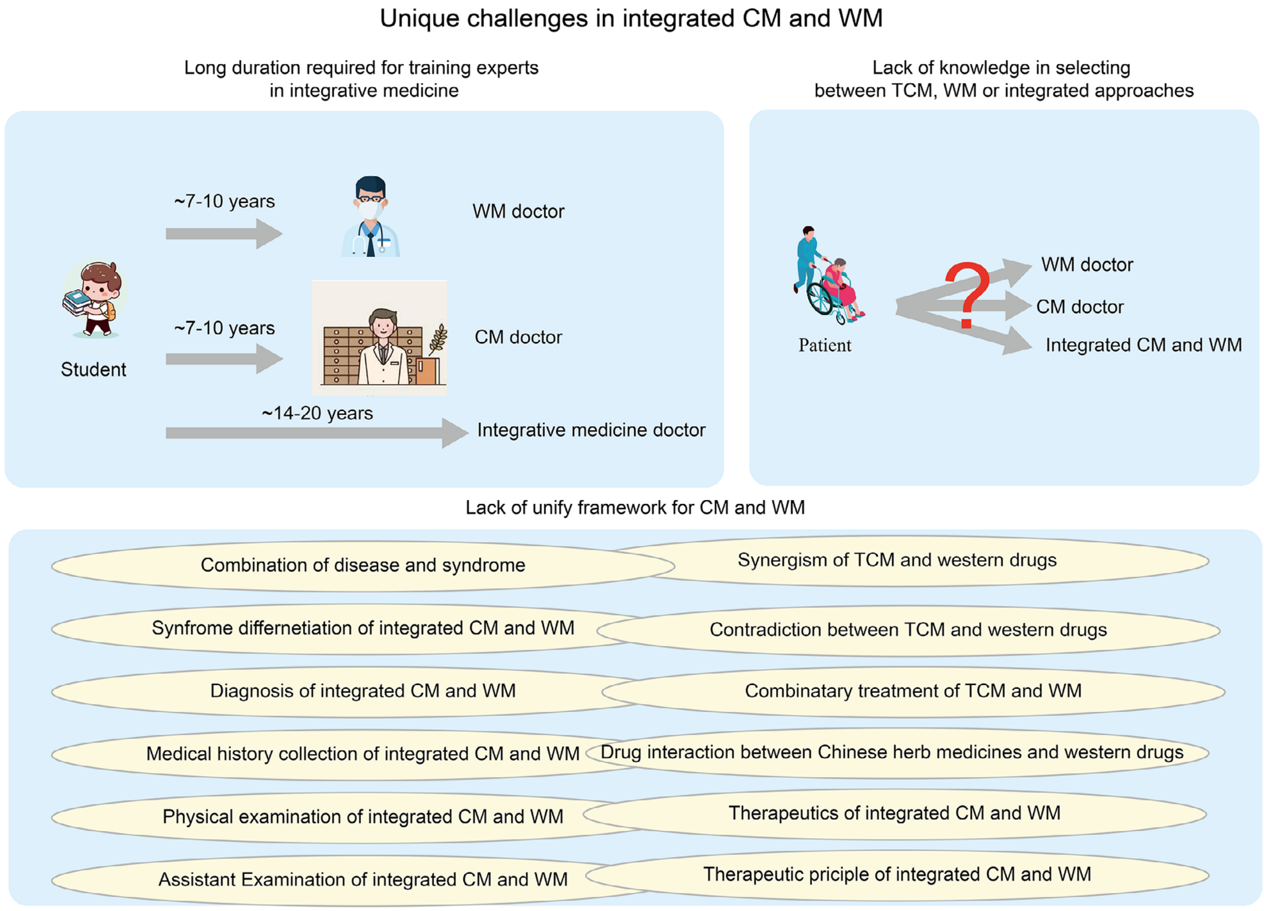


**FIGURE 3** | **Challenges in Integrating Traditional Chinese Medicine and Modern Medicine**.
This figure illustrates the three primary challenges faced in the field of integrative medicine: (1) Extensive training duration needed for students to attain expertise in integrative medicine. (2) Patients often encounter challenges in selecting the most suitable approach for their specific condition due to a lack of knowledge. (3) The absence of a unified framework between Traditional Chinese Medicine and Modern Medicine impedes the development of an integrated approach.

**TABLE 2** | Integrative medicine specific tasks and their difference compared to TCM and Modern Medicine.

| Category | Type | Difference with TCM and Modern Medicine |
|---|---|---|
| Integrative medicine | Combination of disease and syndrome | Focuses on understanding diseases from both biomedical perspectives and TCM syndrome differentiation to provide a holistic diagnosis |
| | Syndrome differentiation of integrated Chinese and Modern Medicine | Utilizes both TCM syndrome patterns and Modern disease classifications to tailor treatments |
| | Diagnosis | Combines Modern diagnostic tools (e.g., lab tests, imaging) with TCM diagnostic methods (e.g., pulse diagnosis, tongue examination) |
| | Medical history and examination | Considers both the Modern medical history collection and physical examinations alongside TCM diagnostic methods |
| | Synergism of TCM and Modern Medicine drugs | Considering the synergistic effects of TCM and Modern Medicine drugs to minimize the effective dosages |
| | Treatment | Integrates TCM therapies (herbal medicine, acupuncture) with Modern treatments (pharmaceuticals, surgeries) to enhance therapeutic outcomes |
| | Therapeutics principle of integrated TCM and Modern Medicine | Aims to provide holistic patient care by combining the strengths of TCM and Modern Medicine therapies |

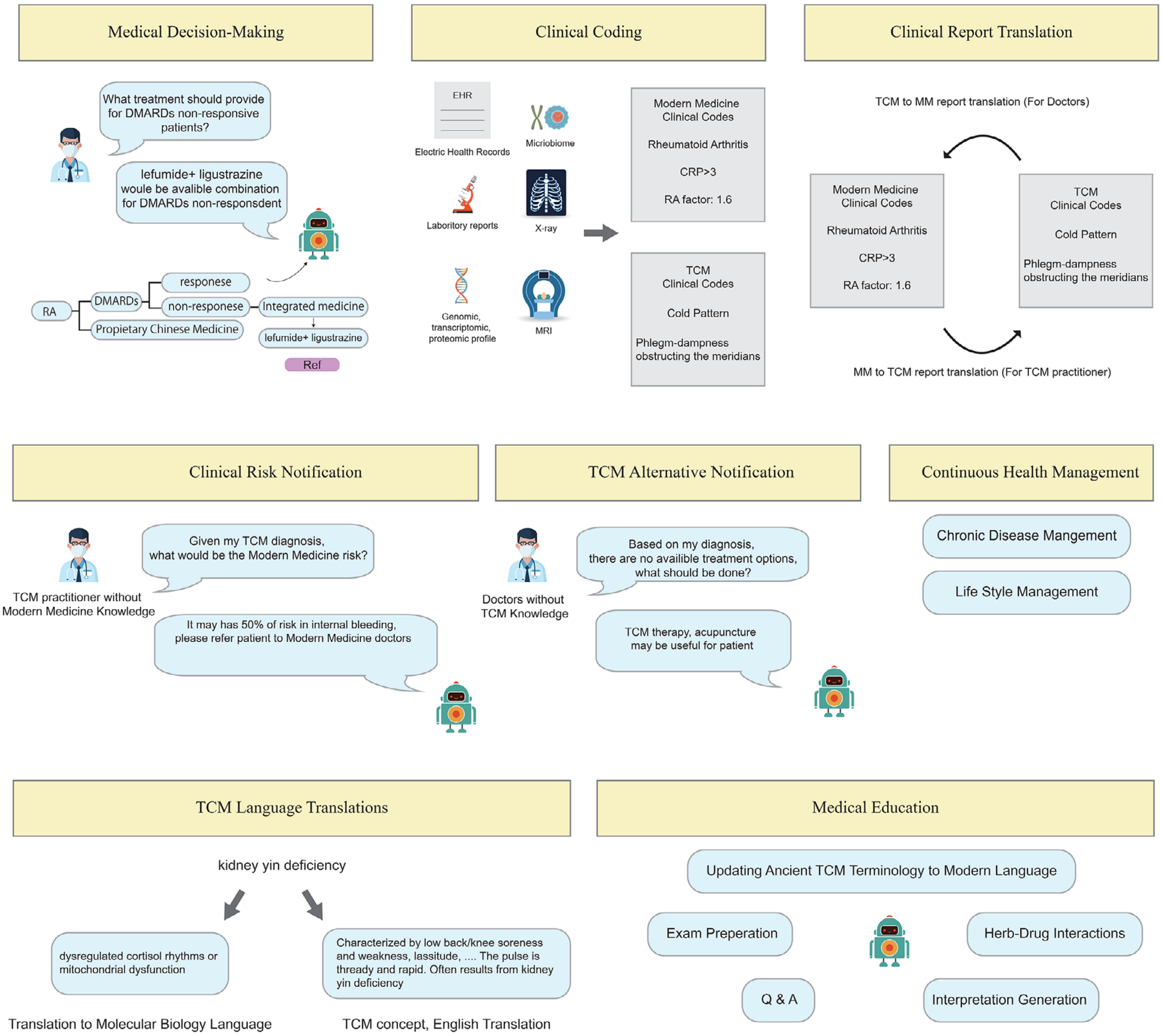


**FIGURE 4** | **Clinical applications for integrative Large Language Models.** Describe the potential applications of Integrative Large Language Models, Medical Decision-Making, Clinical Coding, Clinical Report Translation, Clinical Risk Notification, TCM Alternative Notification, Continuous Health Management, TCM Language Translation, Medical Education.

framework that effectively addresses the complexities of both medical traditions.

## 3 | Opportunities for LLMS in Integrative Medicine

### 3.1 | Clinical Applications

#### 3.1.1 | Clinical Decision-Making

**3.1.1.1 | Medical Decision-Making.** Evidence has shown the effectiveness of Integrative Medicine compared with either TCM/Modern Medicine approach. However, The popularity of integrative medicine has reached a bottleneck due to the limited expertise available in this field. LLMs in Integrative Medicine could help fill this gap while supporting the existing TCM/Modern Medicine practitioners with Integrative Medicine Knowledge (Figure 4). Successful examples in diagnosis using LLMs in TCM/ Modern Medicine. Jiang et al. pioneered using LLMs in diagnosis [74] that used unstructured clinical notes from electronic health to train a clinical language model, achieving 78.7–94.9% AUC in all-purpose clinical predictions. Similarly, TCM LLMs were also developed; HuatuoGPT first landed the basis of the Large Language Model in TCM using LLaMA as a base model and was fine-tuned by TCM question-answer pairs [41]. The remaining task is to deploy a similar training pipeline to fit the scope of Integrative Medicine (Figure 4).

**3.1.1.2 | Clinical Risk Notification.** TCM practitioners in Hong Kong, the United Kingdom, and most countries worldwide have little knowledge of Modern Medicine. However, the emergency risk of some symptoms is high, such as continuous stomach pain, which may indicate internal bleeding in the stomach. Moreover, potential Herb–Drug Interactions, such as the Bai Xian Pi was used in TCM herbal formula, but later, it was reported to induce hepatotoxicity [75, 76]. If the TCM practitioner notices the hepatotoxicity, the herbal formula can be modified according to the patient's health status. It may mitigate the potential danger affecting patients' health (Figure 4).

**3.1.1.3 | TCM Alternative Notification.** More than 7000+ rare diseases without effective drug treatment [77, 78]. Among the diseases with Modern Medicine treatment, some of them are not effective. Among those diseases, TCM has provided an alternative solution. However, due to the existence of modern medical doctors, training seldom includes TCM knowledge. Doctors may not know the alternatives, and integrative LLMs could help to fill the knowledge gap, which improves the patients' health. Integrative LLMs could suggest those cases that provide more patient-centric health care while facing limitations in existing therapy. For example, the TCM Compound (Tongxinluo) in Acute Myocardial Infarction [79], electroacupuncture stimulation on vagal–adrenal axis [80]

### 3.1.2 | Documentation and Coding

**3.1.2.1 | Clinical Coding.** Clinical coding, including systems like the International Classification of Diseases (ICD), medication coding, and procedure coding, is vital in healthcare. It standardizes information about diagnoses, treatments, and procedures, which helps track health data, monitor treatment outcomes, and manage billing and reimbursement.

However, clinical coding is special in TCM, and Modern Medicine is different, so in integrative medicine, clinical coding is different. The curated reports from electric health records, X-rays, microbiomes, laboratory reports, multi-omics, and MRI show that we need simultaneous clinical coding generation for Modern Medicine, and TCM speculations are included (Figure 4).

**3.1.2.2 | Clinical Report Translation.** The application of Integrative Medicine LLMs in clinical report translation bridges the gap between different medical languages and terminologies. By translating TCM concepts into modern medical language, these models enhance the understanding and accessibility of alternative treatments for practitioners trained primarily in Modern Medicine. This capability is crucial for interdisciplinary collaboration, allowing healthcare providers to develop comprehensive treatment plans incorporating diverse therapeutic modalities. Using LLMs in translation promotes more straightforward communication with patients, ensuring they are well-informed about their treatment options and fostering adherence to prescribed therapies (Figure 4).

## 3.2 | Education and Collaboration

### 3.2.1 | TCM Language Translation

TCM has been practiced for thousands of years in China and has developed a unique system of terminology to describe health and disease states within the human body. There are two distinct applications:

**3.2.1.1 | Translating TCM Concept to Molecular Biology Terminology.** TCM concepts are based mainly on holistic and empirical approaches, focusing on patterns of symptoms and overall balance within the body, such as the balance of Yin and Yang or the flow of Qi. These concepts do not easily align with molecular biology's reductionist and mechanistic approach, which focuses on specific biochemical processes and structures.

**3.2.1.2 | Translating TCM Concept Written in Chinese to English.** Many TCM terms do not have direct equivalents in English or Modern medical terminology. This can make it difficult to accurately convey TCM concepts to those unfamiliar with its framework. For instance, terms like "Qi," "Meridians," "Zang-Fu organs," and "Dampness" are unique to TCM and carry specific meanings that are not easily translated without losing some of their original context and nuance.

Efforts to bridge these gaps often involve creating detailed explanations or analogies to help translate TCM concepts into terms that can be understood within the framework of modern science while still respecting traditional knowledge and practices.

### 3.2.2 | Medical Education

In China, TCM and Modern Medicine practitioners must undergo at least 5 years of training. In Hong Kong, the bachelor training programs for TCM and Modern Medicine practitioners span 5 years. Additionally, becoming a specialist in Modern Medicine requires an additional 3–7 years of training, following a structure like that in the United Kingdom. Given the challenges of extensive knowledge required for integrative medicine doctors (expertise in both TCM and Modern Medicine), training integrative medicine experts was long. Based on this, the integrative medicine Large Language Model could be used as a learning assistant for summaries, translations, and explanations.

While medical examinations are essential in TCM and Modern Medicine, detailed explanations accompanying answers are vital for understanding the underlying rationale, a gap that could potentially be filled by leveraging the LLMs to generate comprehensive explanations. A study found that 88.9% of the explanations generated by the Large Language Model contained at least one insightful point [81]. Moreover, it was shown to be able to pass the China National Medical Licensing Examination in both Chinese and English version [82].

Summarizing content from various texts is crucial for successful revision, especially in integrative medicine, which combines knowledge from TCM and MM—two specialized subjects. Therefore, using AI for efficient summarization is vital in aiding effective learning processes. Google NotebookLM has been proposed for AI assistance learning, which helps medical education and utilizes the multimodal adaptability of the backbone Large Language Model Gemini1.5 [83].

Understanding ancient TCM terms can be challenging for both models and users. Therefore, Rui Hua et al. pioneered a methodology where they initially translated ancient texts into modern language and subsequently trained the model based on this updated version [47].

## 3.3 | Continuous Health Management

Healthcare management focuses on maintaining overall health and preventing illnesses. It also involves guiding individuals experiencing symptoms and unsure whether to seek medical advice or the preconsultation stage.

A study conducted in 2017 showed that 97% of participants said that the Internet was often the first medium they used to seek health information [84]. Meanwhile, 71% found the online health information misleading, and 50% of the elderly respondents claimed difficulty searching for information.

LLMs have the potential to summarize vast amounts of online information and filter out misleading information, enhancing accessibility for older adults through conversational interfaces. It can also simplify the process of retrieving information by providing a chat-like experience, facilitating more straightforward access to knowledge for older individuals. Montagna et al. demonstrated the possibility of embedding LLMs in chatbots for chronic disease patient self-management [85]. Mahmood et al. proposed to combine the voice assistant and LLMs into healthcare assistants for older adults [86].

Only a few options like ChatGPT, LLaMA, and Gemini [83] series could be available for maintaining overall health and preventing illnesses, and their response quality was below satisfaction. For TCM LLMs, given the nature of managing suboptimal health conditions, varying dietary recommendations and customized care are based on the personalized constitution.

During the preconsultation phase, a critical juncture influencing patients' decisions on opting for Modern Medicine, Chinese medicine, or an integrated approach, ChatGPT models typically refrain from offering such responses due to their training to steer clear of potentially harmful information. However, fine-tuned models often pivot toward providing direct diagnoses rather than merely recommending the type of expertise individuals should seek out.

For instance, LLMs can suggest dietary adjustments, exercise regimens, and herbal supplements tailored to individual patient profiles. This comprehensive approach addresses immediate health concerns and promotes long-term wellness by encouraging proactive lifestyle changes. By facilitating ongoing monitoring and adjustment of treatment plans, LLMs enhance the overall effectiveness of health management strategies in integrative medicine.

# 4 | Challenges in Integrative Medicine LLMS

## 4.1 | Limited Abilities of Existing LLMs in Integrative Tasks

Existing LLMs in medicine face limitations when addressing Integrative Medicine tasks. Our experiments underscore these limitations across four specific integrative medicine tasks (Figure 5). Further details of the method and evaluation metric can be found in Supplementary 1. These experiments reemphasize the view we proposed in Section 3.4 that LLMs in Integrative Medicine are not an ensemble model that integrates LLMs in Modern Medicine and TCM.

## 4.2 | Technical Challenges

### 4.2.1 | Epistemological Conflict Between TCM and Modern Medicine

TCM/Modern Medicine LLMs act as doctors' assistants, streamlining specific tasks like diagnosis support or treatment recommendations to enhance efficiency. In contrast, Integrative Medicine LLMs serve as doctors' navigators, focusing on bridging the knowledge divide between TCM and Modern Medicine. These systems synthesize insights from both fields, identify potential conflicts (e.g., herb–drug interactions), and optimize holistic care plans. Harmonizing disparate medical paradigms empowers clinicians to deliver safer, more personalized patient care while addressing gaps that single-system models cannot resolve. However, Integrated Medicine has a unique application that unifies TCM and Modern Medicine. The integration of TCM and Western medicine transcends technical compatibility, representing a collision of two distinct epistemologies. While Modern Medicine prioritizes reductionist, biomarker-driven explanations (e.g., targeting specific molecular pathways in hypertension), TCM emphasizes dynamic, systemic imbalances (e.g., "Liver Yang rising" manifesting as migraines or hypertension). LLMs in Integrative Medicine can help mitigate this conflict.

### 4.2.2 | Data Scarcity

According to the Practice Guidelines Registration for Transparency (PREPARE) platform, as of March 5, 2025, there are 3706 clinical practice guidelines, but only 290 pertain to integrative medicine. This number would further decrease when considering the implementability of these guidelines [87]. A critical aspect of integrative medicine is avoiding contradictory treatments between TCM and Modern Medicine, underscoring the need to identify herb–drug interactions. However, the most extensive database for herb–drug interactions, DDID, includes only over 17,000 interactions, despite approximately 25 billion potential herb–drug pairs, given the 9000+ common herbs and 2818 drugs available in DrugBank [88].

### 4.2.3 | Multimodalities in Integrative Medicine

Real-world clinical practices are dynamic decision-making processes, incorporating various diagnostic procedures such as blood tests, imaging studies (X-rays, CT scans, MRIs), omics, metabolites, microbiomes, Electric Health Records, and sensors. Integrating TCM, inspection, listening and smelling, inquiry, and palpation are important aspects of integrative medicine clinic decision-making. However, there are no unifying frameworks to consider the multimodality of integrating TCM multimodality and Modern Medicine multimodality.

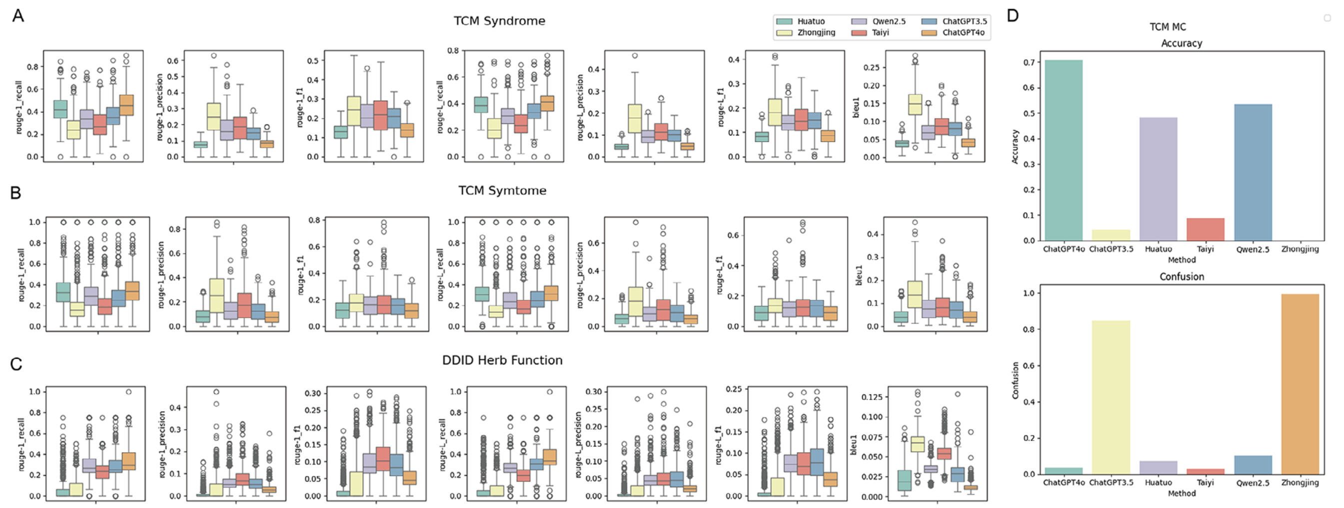


**FIGURE 5** | **Benchmarking the relevance of the Large Language Model's answers in responding to integrative medicine related**. (A) Question and answer of the traditional Chinese medicine syndrome definition. (B) Question and answer of the traditional Chinese medicine symptoms definition. (C) Question and answer of the herb function. (D) Question and answer of the multiple choice, multiple-choice questions related to integrative medicine; "accuracy" indicates that the response matches the correct answer. On the other hand, "confusion" signifies that while the correct answer is included in the response, it is accompanied by other incorrect answers simultaneously.

## 4.3 | Ethical and Legal Barriers

### 4.3.1 | Liability for Errors and Data Privacy

From a legal perspective, although the medicine LLMs tried to improve the accuracy and efficiency of the diagnosis, experience showed that the real-world deployment would not be as good as the in-house experiments. Hence, who should be responsible for the misdiagnosis and even medical malpractice? Gerke et al. first proposed an ethical and legal framework of AI in healthcare [89], which also applied to the Large Language Model. In brief, the four ethical aspects were (1) informed consent to use, (2) safety and transparency, (3) algorithmic fairness and biases, and (4) data privacy. Moreover, the legal aspects were (1) safety and effectiveness, (2) liability, (3) data protection and privacy, (4) cybersecurity, and (5) intellectual property law [89, 90].

### 4.3.2 | Regional Regulatory Compliance

While LLMs strive to deliver the most accurate answers, they frequently overlook whether their responses comply with local regulations. Suggesting drugs that are prohibited in a user's geographical region poses significant risks for patients and can also put clinicians at risk when utilizing these models. LLMs must integrate regulatory considerations into their recommendations to ensure safe and compliant medical guidance.

## 4.4 | Implementation Barriers

### 4.4.1 | Lack of Standardized Operating Procedures

While the accuracy of LLMs is continuously supported by research, no SOP has been developed for LLMs in medicine. In a survey of 174 medical students, 68% were concerned about the liability of AI in medicine, and over 80% agreed that there should be more AI education in medical training [91]. While existing research focuses on improving the accuracy of responses from LLMs in Medicine, there is a significant gap in strategies for their safe implementation and risk mitigation in clinical settings.

# 5 | Proposed Solutions and Future Directions

## 5.1 | Technical Innovations

### 5.1.1 | Real-Time Evidence Retrieval

Updating LLMs frequently is highly impractical due to the intensive resources required for fine-tuning [92]. More practical techniques like Retrieval Augmented Generation (RAG) could help by letting LLMs pull real-time data from medical guidelines or research papers, ensuring responses align with current standards. Mohammad et al. proposed a possible solution that used an adaptive knowledge graph with RAG to incorporate instant knowledge from public databases like PubMed [93].

### 5.1.2 | Improve Response Accuracy

Mixture of Experts (MoE) models could further improve accuracy by assigning subtasks (e.g., diagnosing a TCM pattern vs. analysing a lab result) to specialized "expert" modules. However, most medical LLMs today lack these features, relying instead on oversimplified frameworks. More recently, an evidence-based medicine framework was proposed, and an improvement in medical Q&A was made without modifying the LLMs architecture [94]. Those approaches provide a new perspective on improving the model performance without changing the architecture, but modifying the problem formulations and searching strategy is a cost-effective solution for next-generation LLMs in medicine.

#### 5.1.3 | Improve the Interpretability of the LLMs' response

Existing LLMs would provide the answer to the questions, but few of them would provide a rationale for them to make the decision. It makes it hard to evaluate the quality of the answer. LLMs offer the exciting prospect of in-context few-shot learning via prompting. That is, instead of fine-tuning a separate language model checkpoint for each new task, one can simply "prompt" the model with a few input-output exemplars demonstrating the task. Remarkably, this has been successful for a range of simple question-answering tasks. Wei et al. proposed the concept of chain of thought [56], which prompts a series of intermediate natural language reasoning steps that lead to the final output. Despite their success, LLMs are still limited by a lack of knowledge and are prone to hallucinations during reasoning, which can lead to errors in reasoning processes. Recently, Luo et al. [95] proposed combining the knowledge graph with reasoning since the KGs capture abundant factual knowledge in a structured format, providing a faithful reasoning source. However, there is a need to ensure the quality of the data in the knowledge graph. Yang et al. demonstrated that malicious links to the drug within a knowledge graph (KG) could contaminate the LLMs responses generated based on the information stored in the KG [96]. In the future, improving the human interpretability of the LLMs would be a promising direction, allowing doctors and patients to trust the model output and notice if mistakes are generated from the model. Also, the Large Language Model could inspire us to notice some hidden patterns humans have not noticed.

#### 5.1.4 | Merging Subjective TCM Modality Into Multimodal Frameworks

The unique characteristics of multimodality in Integrative Medicine have been identified. Some existing work could serve as a reference for modality integration. For example, Belyaeva et al. proposed a multimodal LLM named HeLM, which can effectively use demographic and clinical features in addition to high-dimensional time-series data to estimate disease risk [97]. Xu et al. proposed that ELSXR align the X-ray image with its paired report to improve visual question and answering (VQA) and question and answering (Q&A) [98]. Considering the integration of syndrome (i.e., mainly from TCM diagnosis, which is relatively subjective) with the objective data obtained from these diagnostic modalities is essential for constructing a holistic view of the patient's health status. We proposed to avoid the syndrome classification at the beginning, and instead, we should use the tongue image, patients' pulse signal profile, and the saliva metabolic profile as the input for the Multimodality integration of TCM and Modern Medicine. After that, the linkage with the syndrome could be established, such as tongue patterns associated with TCM syndrome [99] polycystic ovary syndrome connected with pulse wave [100].

### 5.2 | Data and Standardization

#### 5.2.1 | Curating Data Sources for Integrative Medicine

**5.2.1.1 | Data From ICD-11 TCM Cohort and WHO.** Translation involves rendering Chinese terms into English and updating ancient TCM terminology to modern language. The 11th edition of the International Statistical Classification of Diseases and Related Health Problems (ICD-11) incorporated TCM terminology in 2019 [101]. This integration acknowledges TCM's significance in the global health system and is pivotal in translating TCM concepts into English. The integration enables Modern Medicine practitioners and healthcare professionals to grasp the essence and influence of TCM on patient care. Also, WHO International Standard terminologies on TCM provided a more comprehensive list of the standard translation of TCM concepts to English [102].

**5.2.1.2 | Data From Chinese Proprietary Drug.** From the perspective of establishing integrative medicine LLMs to optimize treatment efficacy through combination therapies, CPMs offer three key advantages: (1) Unlike TCM complex syndrome differentiation system, CPMs adopt a disease- and symptom-oriented framework, for example, the clinical recommendation of Tripterygium glycosides with csDMARDs for rheumatoid arthritis patients is not necessary alongside with syndrome differentiation [103, 104]. This standardized approach bridges the epistemological conflict between TCM and Modern Medicine, which helps the LLMs establish a cohesive framework. (2) CPMs' fixed ingredient ratios and quantified production standards enable LLMs to effectively utilize precise chemical composition databases alongside modern pharmacokinetic models. This integration allows for more accurate predictions of dose-dependent interactions in Herb–Drug combinations [105], significantly enhancing safety assessments in polypharmacy regimens. (3) The prescription stability of CPMs supports their use as reproducible intervention units. LLMs can use CPMs to identify the related multicenter Randomized Control Trials, while traditional herbal prescriptions could not be due to the high variation among different trails.

**5.2.1.3 | Data from EBM.** The integration of EBM offers a scaffold to address epistemological conflicts in cross-paradigm care. For instance, in DMARDs-resistant cases, EBM ensures therapeutic choices remain rooted in clinical trial data. At the same time, MB decodes how TCM herbs like demethylzeylasteral (an ingredient in Tripterygium wilfordii) modulate JAK-STAT pathways [106], translating ancient "damp-heat" syndrome concepts into actionable anti-inflammatory targets. Systems Medicine contextualizes these insights, modelling interactions between herbal compounds, gut microbiota, and immune cascades to predict holistic outcomes.

**5.2.1.4 | Data From Integrative Medicine Research.** We should utilize the large corpus of TCM research in China. Cheng et al. reviewed the biological basis of TCM syndrome and proposed the five-step principles for TCM syndrome standardization [107]—the proteomic and gene expression characteristic of "blood stasis" [108, 109]. Ya et al. decoded the metabolic characteristics of cold hot pattern within RA patients [110]. Later, the hot individuals were further summarized as related to inflammation, immune-related pathways, and higher metabolism. Meanwhile, lower metabolism has been observed in cold people [111].

**5.2.1.5 | Data From Molecular Biology and Systems Medicine.** It is essential to incorporate a larger corpus of research from molecular biology and systems medicine, as curated in databases like PubMed and Google Scholar. Leveraging these sources allows LLMs to summarize research findings and synthesize clinical practice guidelines effectively. Additionally, optimizing existing drug-drug interaction databases and resources like TCMbank is crucial. TCMbank can be a valuable resource for LLMs to learn about herb–drug interactions, as it curates comprehensive chemical structures of ingredients [105]. By transforming herb–drug interactions into drug-drug interaction problems, we can decompose herbs into multiple ingredient–drug pairs and predict their interactions independently. This approach enhances the integration of TCM and Modern Medicine practices, promoting safer and more effective clinical guidelines in integrative medicine.

#### 5.2.2 | Bridging Epidemiological Conflicts Through TCM Standardization

We should utilize the large corpus of TCM research in China. Cheng et al. reviewed the biological basis of TCM syndrome and proposed the five-step principle for TCM syndrome standardization [107], which identified the proteomic and gene expression characteristic of "blood stasis" [108, 109]. Ya et al. decoded the metabolic characteristics of Cold-Hot patterns within RA patients [110]. Later, the Hot individuals were further summarized as related to inflammation, immune-related pathways, and higher metabolism. Meanwhile, lower metabolism has been observed in Cold people [111].

### 5.3 | Policy and Education

#### 5.3.1 | Revealing the Effect of Regional Regulation

As a result, the LLMs in practice for practice needed to be curated with the list of authorized Chinese Proprietary Medicine (CPMs), drugs, and sponsors drugs. For example, the list of authorized CPMs is available at https://data.gov.hk/tc-data/dataset/hk-dh-cmd-cmd-list-of-proprietary-chinese-medicine, and the list of available drugs under China Healthcare Security could be found in http://big5.www.gov.cn/gate/big5/www.gov.cn/zhengce/zhengceku/202411/content_6989859.htm. In the future, the response of LLM in Integrative Medicine should be customized based on the region and patients' financial ability.

#### 5.3.2 | Establish SOP for Clinical Practitioners

Should an SOP be provided to clinicians to tell them the standard prompt for communicating with LLMs? When to believe the LLMs response? When should they further ask the LLMs to validate its response? The potential framework of SOP in LLMs for clinicians should be established to achieve safe and effective use in clinics. It should include the process, objective, responsibility, procedure, and the standard of procedures.

### 5.4 | Three-Stage Development Strategy

#### 5.4.1 | Stage 1 Evidence Summarization

In this stage, LLM's capabilities include providing evidence-based responses, where the model answers a question and explains the reasoning behind its answer (Figure 6). Training strategy Integrative medicine combines knowledge from Modern Medicine and TCM. Since much of the TCM context is Chinese, the chosen base model must comprehend Chinese and English. Subsequently, training in integrative medicine content is necessary. This field involves domain-specific knowledge from integrative medicine textbooks, clinical guidelines, and trials. Once the integrative medicine corpus is curated in text format, it can further pretrain the base model. The goal is to embed domain-specific knowledge into the model. Following pretraining, an instruction-tuning dataset can be constructed based on the corpus. This dataset typically consists of question-answer pairs related to integrative medicine. While the model may accurately answer integrative medicine questions post-training, it may lack the ability to justify its deductions. RAG can enhance LLMs interpretability [112]. RAG ensures that the generated answers are accompanied by references that validate their accuracy, like Pre-Retrieval Process [113] and Post-Retrieval Process like LangChain [114]. Even more, some would employ MoE models, which input the query to multiple experts independently, and the answers were weighted according to the gate network weighting [115, 116].

#### 5.4.2 | Stage 2 Hypothetical Evidence Generation

LLMs progress to generating hypothetical answers (Figure 6). The focus of stage 2 is given the fact that only limited numbers of integrative medicine clinical practice guidelines [117]. These answers may not be explicitly stated in the source material but are implied or inferred. To incorporate the ability to generate in LLMs, we needed to embed the fingerprint of the molecules into the model. Given the fine-tuned-based model trained on stage 1, we must fine-tune it further. It uses the human preference data to avoid the model's output of toxic information and respond more accurately. Also, molecule features needed to be integrated into the model, from structure to targets, cell response, network and clinical information. Such data could allow our LLM to understand the linking between drug features and the disease. Moreover, it induces similar molecules that could target the same disease. Huang et al. developed the first Large Language Model that could predict drug repurposing [77]. Our perspective is to further extent the model's applicability to arbitrary molecules, even the TCM-derived compounds.

#### 5.4.3 | Stage 3 Integration of Patients' multimodal Data for Precision Diagnosis

Moving forward, LLMs evolve to integrate patients' diverse data types for precise diagnosis (Figure 6). While Modern Medicine typically focuses on diseases, the reality is that patients are present with a combination of conditions. By considering the specificities of each patient, the model aims to reduce the chances of treatment non-responsiveness and enhance personalized

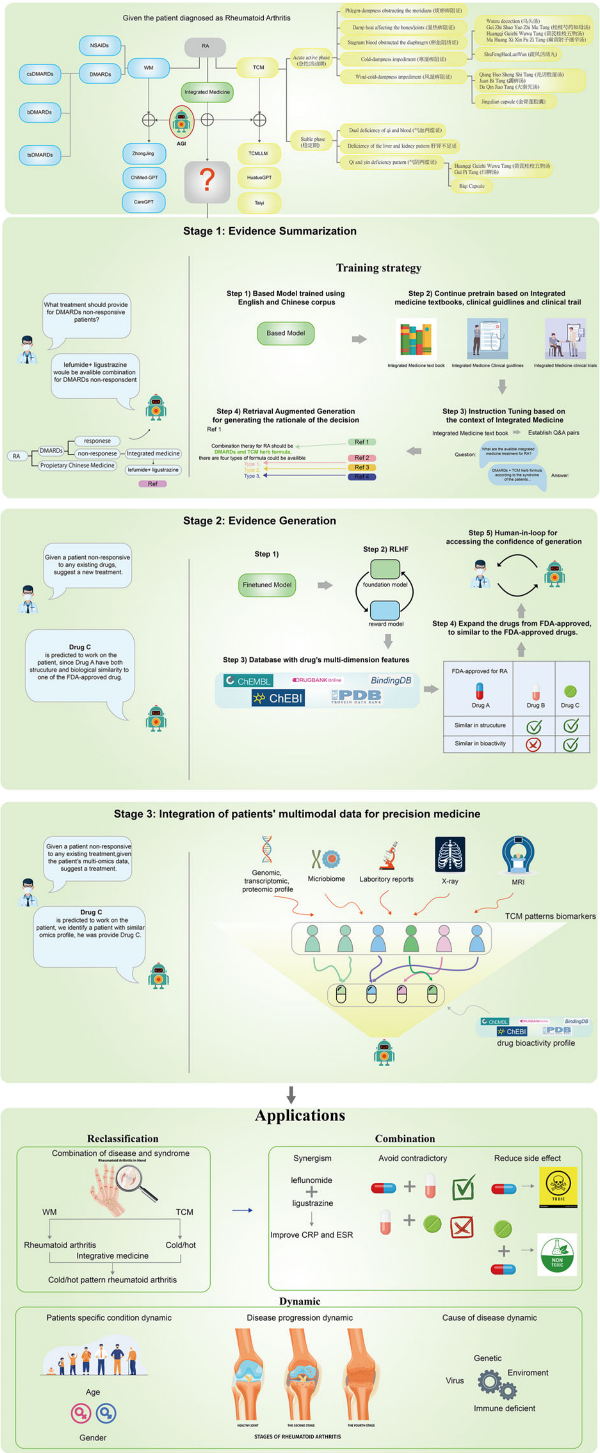


**FIGURE 6** | **Strategy to develop integrative medicine Large Language Model**.
The gap in integrative medicine LLM, the proposed three stages strategies to train an integrative medicine LLM, and its application.

healthcare outcomes. In the final stage of the process, patient-specific data is integrated into the model to identify key disease-causing factors and accurately classify patients based on biomarkers and TCM pattern biomarkers. The patient data, being multimodal, necessitates the deployment of individual modality models—such as visual models for X-rays, MRIs and histopathology imaging [118, 119]. Each modality model is designed to capture the nuances specific to that data type. The algorithm is structured to map patient features to multi-dimensional drug features. The concept of CMAP could provide the connectivity score for this mapping. Subsequently, the insights gathered from each modality are consolidated using LLMs. This comprehensive approach enables precise decision-making tailored to each patient's condition, ensuring targeted treatment strategies based on a holistic understanding of the patient's health profile.

## 6 | Discussion

Integrating LLMs into integrative medicine marks a pivotal advancement in harmonizing the empirical, systemic approaches of TCM with the evidence-based precision of modern medical science. LLMs hold immense potential to bridge these paradigms, enabling cross-disciplinary knowledge synthesis, clinical decision support, and personalized care. Their deployment faces multifaceted challenges rooted in epistemological, technical, and ethical complexities. Addressing these challenges requires innovative solutions anchored in interdisciplinary collaboration and rigorous validation.

A core barrier lies in reconciling TCM's holistic frameworks, such as Qi imbalances or Liver Yang rising, with Modern Medicine's reductionist focus on molecular pathways. LLMs must transcend mere terminology translation to establish mechanistic links between these systems. For instance, mapping TCM syndromes like blood stasis to proteomic biomarkers [108, 109] or connecting rheumatoid arthritis-associated cold-hot patterns to distinct metabolic profiles [110] could operationalize TCM diagnostics for modern clinical workflows. Hybrid models like Taiyi [38], which align bilingual TCM and Modern Medicine corpus, demonstrate this potential but require scaling to address the scarcity of high-quality integrative datasets.

Technically, existing LLMs struggle with integrative tasks due to fragmented training data and monolingual biases. Architectures like MoE could allocate specialized modules, such as a TCM syndrome differentiation expert trained on pulse/tongue data and a modern diagnostics module analysing lab results and emulating clinician workflows. Innovations like BianQue's Chain of Questioning [40], which refines diagnoses through iterative inquiry, offer a template for dynamic reasoning. However, computational costs remain prohibitive; parameter-efficient fine-tuning (e.g., LoRA in ShenNong-TCM-LLM) and multimodal integration of wearable devices or omics data may mitigate these barriers. LLMs must dynamically update knowledge to avoid perpetuating outdated practices from historical TCM texts, such as using toxic compounds like cinnabar. RAG systems tied to real-time PubMed/ICD-11 feeds could ensure alignment with evolving standards.

Clinically, LLMs could democratize integrative expertise by alerting TCM practitioners to modern medical risks (e.g., Bai Xian Pi hepatotoxicity) or suggesting evidence-based alternatives for refractory conditions [75, 76]. However, ethical and regulatory frameworks are critical to mitigate risks. Embedding region-specific formularies (e.g., China's approved CPM list) into LLMs can prevent recommendations of unapproved therapies, while transparency protocols, such as disclosing model limitations and requiring clinician validation of AI-generated diagnoses. Standardized workflows, informed by frameworks like Gerke et al. pillars of AI ethics, must balance innovation with patient safety [89].

The three-stage pipeline for establishing LLMs in integrative medicine was proposed to fill the gaps between state-of-the-art LLMs in tasks of integrative medicine, which aims to fit the needs of clinics. Global adaptability will require leveraging WHO's ICD-11 TCM terminology standardization [101]. Patient-centric tools, such as LLM-powered chatbots for personalized dietary advice (e.g., constitution-based TCM regimens), could further enhance adherence to integrative care.

Despite this promise, limitations persist. Untranslatable TCM concepts (e.g., Qi) risk misinterpretation without robust cultural-linguistic safeguards and siloed validation in controlled settings limits real-world applicability. Computational costs and subjective concepts of TCM syndrome further constrain progress. Evidence-Based Medicine illustrated the transformative potential of bridging these gaps, for example, a TCM compound Tongxinluo was validated in acute myocardial infarction trials [79].

In conclusion, LLMs in integrative medicine stand at the frontier of a healthcare revolution, poised to harmonize ancient wisdom with cutting-edge science. Success hinges on collaborative innovation, uniting AI developers, clinicians, and policymakers to refine architectures, enforce ethical rigor, and validate tools in diverse clinical landscapes. By doing so, these models could redefine global standards for personalized, holistic care, ensuring AI is a collaborator rather than a disruptor in the art and science of healing.

**Acknowledgments**

This work is supported by the Technology Innovation Strategy Special Fund (Guangdong-Hong Kong-Macau Joint Lab, 2020B1212030006), the HK Theme-based Scheme (T12-201/20-R), Hong Kong General Research Fund (12102722), Young Collaborative Research Grant (RGC/C2004-23Y), and Guangdong-Hong Kong Technology Cooperation Funding Scheme (GHX/270/22SZ).

**Conflicts of Interest**

The authors declare no conflicts of interest.

**References**

1. H. Xu, Y. Zhang, P. Wang, et al., "A Comprehensive Review of Integrative Pharmacology-Based Investigation: A Paradigm Shift in Traditional Chinese Medicine," *Acta Pharmaceutica Sinica B* 11, no. 6 (2021): 1379–1399.

2. W.-J. Wang and T. Zhang, "Integration of Traditional Chinese Medicine and Western Medicine in the Era of Precision Medicine," *Journal of Integrative Medicine* 15, no. 1 (2017): 1–7.

3. L. Y. Kong and R. X. Tan, "Artemisinin, a Miracle of Traditional Chinese Medicine," *Natural Product Reports* 32, no. 12 (2015): 1617–1621.

4. J. Fleckenstein, D. Irnich, N. Goldman, et al., "Adenosine A1 Receptors Mediate Local Anti-nociceptive Effects of Acupuncture," *Deutsche Zeitschrift Für Akupunktur* 53, no. 3 (2010): 38–39.

5. G. K. K. Leung, S. W. H. Wong, G. K. B. Ng, and K. N. Hung, "Concomitant Use of Western and Chinese Medicine Treatments in Neurosurgical Patients in Hong Kong," *Chinese Journal of Integrative Medicine* (2011), https://doi.org/10.1007/s11655-011-0818-8.

6. OpenAI, J. Achiam, S. Adler, S. Agarwal et al., eds. GPT-4 Technical Report (2024), https://arxiv.org/abs/2303.08774.

7. K. R. Laukamp, R. A. Terzis, J. M. Werner, et al., "Monitoring Patients With Glioblastoma by Using a Large Language Model: Accurate Summarization of Radiology Reports With GPT-4," *Radiology* 312, no. 1 (2024): e232640.

8. S. Li and B. Zhang, "Traditional Chinese Medicine Network Pharmacology: Theory, Methodology and Application," *Chinese Journal of Natural Medicines* 11, no. 2 (2013): 110–120.

9. X. Li, G. Yang, X. Li, et al., "Traditional Chinese Medicine in Cancer Care: A Review of Controlled Clinical Studies Published in Chinese," *PLoS ONE* 8, no. 4 (2013): e60338.

10. L. Lu, S. Ni, X. He, Y. Huang, X. Chen, and Z. Yang, "From Tradition to Evidence-Base: Leveraging TCM Human Use Experience in Modern Drug Development," *Pharmacological Research—Modern Chinese Medicine* 13 (2024): 100535.

11. A. Radford and K. Narasimhan, eds., Improving Language Understanding by Generative Pre-Training (2018).

12. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, eds., Language Models Are Unsupervised Multitask Learners (2019).

13. H. Touvron, L. Martin, K. R. Stone, et al., "Llama 2: Open Foundation and Fine-Tuned Chat Models," *ArXiv* (2023). abs/2307.09288.

14. A. Dubey, A. Jauhri, A. Pandey, et al., "The Llama 3 Herd of Models," *ArXiv* (2024). abs/2407.21783.

15. H. Touvron, T. Lavril, G. Izacard, et al., "LLaMA: Open and Efficient Foundation Language Models," *ArXiv* (2023). abs/2302.13971.

16. A. Yang, B. Yang, B. Hui, et al., "Qwen2 Technical Report," *ArXiv* (2024). abs/2407.10671.

17. R. Gan, Z. Wu, R. Sun, et al., "Ziya2: Data-Centric Learning Is All LLMs Need," *ArXiv* (2023). abs/2311.03301.

18. T. G. A. Zeng, B. Xu, B. Wang, et al., "ChatGLM: A Family of Large Language Models From GLM-130B to GLM-4 all Tools," *ArXiv* (2024). abs/2406.12793.

19. T. L. Scao, A. Fan, C. Akiki, E. Pavlick, S. Ili'c, D. Hesslow, et al., "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model," *ArXiv* (2022). abs/2211.05100.

20. A. M. Yang, B. Xiao, B. Wang, et al., "Baichuan 2: Open Large-Scale Language Models," *ArXiv* (2023). abs/2309.10305.

21. R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn, "Direct Preference Optimization: Your Language Model Is Secretly a Reward Model," In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA: Curran Associates Inc., 2024): 2338.

22. E. Ullah, A. Parwani, M. M. Baig, and R. Singh, "Challenges and Barriers of Using Large Language Models (LLM) Such as ChatGPT for Diagnostic Medicine With a Focus on Digital Pathology—A Recent Scoping Review," *Diagnostic Pathology* 19, no. 1 (2024): 43.

23. P. K. Kanithi, C. Ce, M. A. Pimentel, et al., "MEDIC: Towards a Comprehensive Framework for Evaluating LLMs in Clinical Applications," *ArXiv* (2024). abs/2409.07314.

24. K. Singhal, S. Azizi, T. Tu, et al., "Large Language Models Encode Clinical Knowledge," *Nature* 620 (2022): 172–180.

25. J. Wang, Y. Zhang, L. Zhang, et al., "Fengshenbang 1.0: Being the Foundation of Chinese Cognitive Intelligence," *ArXiv* (2022). abs/2209.02970.

26. J. Bai, S. Bai, Y. Chu, et al., "Qwen Technical Report," *ArXiv* (2023). abs/2309.16609.

27. A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, and A. N. Gomez, "Attention Is All You Need," In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates Inc., 2017): 6000–6010.

28. Z. Du, Y. Qian, X. Liu, M. Ding, J. Qiu, Z. Yang, and J. Tang, "GLM: General Language Model Pretraining With Autoregressive Blank Infilling," In *Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, 2022): 320–335.

29. Y. Cui, Z. Yang, and X. Yao, "Efficient and Effective Text Encoding for Chinese LLaMA and Alpaca," *ArXiv* (2023). abs/2304.08177.

30. M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, and B. Catanzaro, "Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism," *ArXiv* (2019). abs/1909.08053.

31. J. Li, D. Li, S. Savarese, and S. C. H. Hoi, "BLIP-2: Bootstrapping Language-Image Pre-training With Frozen Image Encoders and Large Language Models," In *International Conference on Machine Learning* (JMLR.org, 2023).

32. M. Ding, Z. Yang, W. Hong, W. Zheng, C. Zhou, and D. Yin, "CogView: Mastering Text-to-Image Generation via Transformers," In *Neural Information Processing Systems* (Curran Associates, Inc., 2021): 19822–19835, https://proceedings.neurips.cc/paper_files/paper/2021/file/a4d92e2cd541fca87e4620aba658316d-Paper.pdf.

33. T.-Y. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, and D. Ramanan, "Microsoft COCO: Common Objects in Context," In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds). *Computer Vision – ECCV 2014*. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham., 2014), https://doi.org/10.1007/978-3-319-10602-1_48.

34. P. Sharma, N. Ding, S. Goodman, and R. Soricut, "Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset for Automatic Image Captioning," In: Gurevych, I., Miyao, Y., (eds). *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (Association for Computational Linguistics, 2018): 2556–2565, https://aclanthology.org/P18-1238/.

35. D. Demner-Fushman, M. D. Kohli, M. B. Rosenman, et al., "Preparing a Collection of Radiology Examinations for Distribution and Retrieval," *Journal of the American Medical Informatics Association: JAMIA* 23, no. 2 (2015): 304–310.

36. J. Chen, H. Guo, K. Yi, B. A. Li, and M Elhoseiny, "VisualGPT: Data-Efficient Adaptation of Pretrained Language Models for Image Captioning," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022): 18030–18040.

37. Y. Tian, R. Gan, Y. Song, J. Zhang, and Y. Zhang, eds., *ChiMed-GPT: A Chinese Medical Large Language Model With Full Training Regime and Better Alignment to Human Preferences* (Bangkok, Thailand: Association for Computational Linguistics, 2024 August).

38. L. Luo, J. Ning, Y. Zhao, et al., "Taiyi: A Bilingual Fine-Tuned Large Language Model for Diverse Biomedical Tasks," *Journal of the American Medical Informatics Association: JAMIA* 31, no. 9 (2024): 1865–1874, https://doi.org/10.1093/jamia/ocae037.

39. S. Yang, H. Zhao, S. Zhu, G. Zhou, H. Xu, Y. Jia, and H. Zan, "Zhongjing: Enhancing the Chinese Medical Capabilities of Large Language Model Through Expert Feedback and Real-World Multi-Turn Dialogue," In *AAAI Conference on Artificial Intelligence* (AAAI Press, 2024), https://doi.org/10.1609/aaai.v38i17.29907.

40. Y. Chen, Z. Wang, X. Xing, et al., "BianQue: Balancing the Questioning and Suggestion Ability of Health LLMs With Multi-Turn Health Conversations Polished by ChatGPT," *ArXiv* (2023). abs/2310.15896.

41. H. Zhang, J. Chen, F. Jiang, F. Yu, Z. Chen, and J. Li, "HuatuoGPT, Towards Taming Language Model to Be a Doctor," In: Bouamor, H., Pino, J., Bali, K., (eds). *Findings of the Association for Computational Linguistics: EMNLP 2023* (Association for Computational Linguistics, 2023): 10859–10885, https://aclanthology.org/2023.findings-emnlp.725/.

42. W. Y. Wei Zhu and X. Wang, *ShenNong-TCM: A Traditional Chinese Medicine Large Language Model* (GitHub, 2023).

43. W. Rongsheng, *CareGPT: Medical LLM, Open Source Driven for a Healthy Future* (GitHub, 2023).

44. Z. Bao, W. Chen, S. Xiao, K. Ren, J. Wu, C. Zhong, et al., "DISC-MedLLM: Bridging General Large Language Models and Real-World Medical Consultation," *ArXiv* (2023). abs/2308.14346.

45. R. Wang, Y. Duan, C. Lam, J. Chen, J. Xu, and H. Chen, "IvyGPT: InteractiVe Chinese Pathway Language Model in Medical Domain," *Artificial Intelligence. CICAI 2023. Lecture Notes in Computer Science* (Singapore: Springer Nature Singapore; 2024).

46. X. Zhou, K. Yang, H. Tian, X. Dong, K. Xu, and R. Hua. TCMLLM GitHub; 2023, https://github.com/2020MEAI/TCMLLM.

47. R. Hua, X. Dong, Y. Wei, et al., "Lingdan: Enhancing Encoding of Traditional Chinese Medicine Knowledge for Clinical Reasoning Tasks With Large Language Models," *Journal of the American Medical Informatics Association* 31, no. 9 (2024): 2019–2029.

48. T. Tu, A. Palepu, M. Schaekermann, et al., "Towards Conversational Diagnostic AI," *ArXiv* (2024). abs/2401.05654.

49. K. Saab, T. Tu, W.-H. Weng, et al., "Capabilities of Gemini Models in Medicine," *ArXiv* (2024). abs/2404.18416.

50. C. Wu, W. Lin, X. Zhang, Y. Zhang, W. Xie, and Y. Wang, "PMC-LLaMA: Toward Building Open-Source Language Models for Medicine," *Journal of the American Medical Informatics Association* 31, no. 9 (2024): 1833–1843.

51. E. Bolton, A. Venigalla, M. Yasunaga, et al., "BioMedLM: A 2.7B Parameter Language Model Trained on Biomedical Text," *ArXiv* (2024). abs/2403.18421.

52. T. J. Pollard, A. E. W. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, "The eICU Collaborative Research Database, a Freely Available Multi-Center Database for Critical Care Research," *Scientific Data* 5 (2018).

53. A. E. W. Johnson, L. Bulgarelli, L. Shen, et al., "MIMIC-IV, a Freely Accessible Electronic Health Record Dataset," *Scientific Data* 10 (2023).

54. O. Ben Shoham and N. Rappoport, "CPLLM: Clinical Prediction With Large Language Models," *PLOS Digital Health* 3, no. 12 (2024): e0000680.

55. C. Ce, P. K. Kanithi, P. Munjal, et al., "Med42 - Evaluating Fine-Tuning Strategies for Medical LLMs: Full-Parameter vs. Parameter-Efficient Approaches," *ArXiv* (2024). abs/2404.14779.

56. T. B. Brown, B. Mann, N. Ryder, et al., "Language Models Are Few-Shot Learners," in *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada: Curran Associates Inc.; 2020): 159.

57. M. Nye, A. Andreassen, G. Gur-Ari, et al., "Show Your Work: Scratchpads for Intermediate Computation With Language Models," *ArXiv* (2021). abs/2112.00114.

58. J. Wei, X. Wang, D. Schuurmans, et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (Red Hook, NY, USA: Curran Associates Inc.; 2022).

59. D. Zhou, N. Scharli, L. Hou, et al., "Least-to-Most Prompting Enables Complex Reasoning in Large Language Models," *ArXiv* (2022). abs/2205.10625.

60. D. Jin, E. Pan, N. Oufattole, W.-H. Weng, H. Fang, and P. Szolovits, "What Disease Does This Patient Have? A Large-Scale Open Domain Question Answering Dataset From Medical Exams," *Applied Sciences* 11 (2021): 6421.

61. Pal A, Umapathi LK, Sankarasubbu M, eds., *MedMCQA: A Large-Scale Multi-Subject Multi-Choice Dataset for Medical Domain Question Answering* (ACM Conference on Health, Inference, and Learning, 2022).

62. G. Yang, X. Liu, J. Shi, Z. Wang, and G. Wang, "TCM-GPT: Efficient Pre-training of Large Language Models for Domain Adaptation

in Traditional Chinese Medicine," *Computer Methods and Programs in Biomedicine Update* 6 (2024): 100158.

63. O. Chin, N. S. Jamil, Z. Zainudin, N. A. Hitam, N. Ibrahim, and A. H. Ahmad Sa'ahiry, "OYEN: A User-Centric LLM-Based Bilingual Healthcare Chatbot," in *2024 5th International Conference on Artificial Intelligence and Data Sciences (AiDAS)* (2024): 435–440, https://doi.org/10.1109/AiDAS63860.2024.10730140.

64. H. Zhang, X. Wang, Z. Meng, Y. Jia, and D. Xu, "Qibo: A Large Language Model for Traditional Chinese Medicine," *ArXiv* (2024). abs/2403.16056.

65. S. Wei, X. Peng, Y.-f Wang, et al., "BianCang: A Traditional Chinese Medicine Large Language Model," *ArXiv* (2024). abs/2411.11027.

66. Y. Li, Z. Li, K. Zhang, R. Dan, S. Jiang, and Z. Y. ChatDoctor, "A Medical Chat Model Fine-Tuned on a Large Language Model Meta-AI (LLaMA) Using Medical Domain Knowledge," *Cureus* 15, no. 6 (2023): e40895, https://doi.org/10.7759/cureus.40895.

67. T. Han, L. C. Adams, J.-M. Papaioannou, et al., "MedAlpaca—An Open-Source Collection of Medical Conversational AI Models and Training Data," *ArXiv* (2023). abs/2304.08247.

68. H. Xiong, S. Wang, Y. Zhu, et al., "DoctorGLM: Fine-tuning Your Chinese Doctor Is Not a Herculean Task," *ArXiv* (2023). abs/2304.01097.

69. M. Xu, "MedicalGPT: Training Medical GPT Model," 2023, https://github.com/shibing624/MedicalGPT.

70. P. Densen, "Challenges and Opportunities Facing Medical Education," *Transactions of the American Clinical and Climatological Association* 122 (2011): 48–58.

71. Q. C. Thawakar, A. M. Shaker, S. S. Mullappilly, H. Cholakkal, R. M. Anwer, and S. Khan, eds., *XrayGPT: Chest Radiographs Summarization Using Large Medical Vision-Language Models* (Bangkok, Thailand: Association for Computational Linguistics, 2024 August).

72. M. Tran, P. Schmidle, S. J. Wagner, et al., "Generating Highly Accurate Pathology Reports From Gigapixel Whole Slide Images With HistoGPT," *Medrxiv* (2024). 2024.03.15.24304211.

73. D. Van Veen, C. Van Uden, L. Blankemeier, et al., "Adapted Large Language Models Can Outperform Medical Experts in Clinical Text Summarization," *Nature Medicine* 30, no. 4 (2024): 1134–1142.

74. L. Y. Jiang, X. C. Liu, N. P. Nejatian, et al., "Health System-Scale Language Models Are All-Purpose Prediction Engines," *Nature* 619, no. 7969 (2023): 357–362.

75. L. Perharic-Walton and V. Murray, "Toxicity of Chinese Herbal Remedies," *Lancet* 340, no. 8820 (1992): 674.

76. S. H. Kang, J. I. Kim, K. H. Jeong, et al., "Clinical Characteristics of 159 Cases of Acute Toxic Hepatitis," *The Korean Journal of Hepatology* 14, no. 4 (2008): 483–492.

77. K. Huang, P. Chandak, Q. Wang, et al., "A Foundation Model for Clinician-Centered Drug Repurposing," *Nature Medicine* 30 (2024): 3601–3613, https://doi.org/10.1038/s41591-024-03233-x.

78. R. Facharztmagazine, "Rare Disease Day 2021: Familiäres Mittelmeerfieber," *Orthopädie & Rheuma* 24 (2021): 62.

79. Y. Yang, X. Li, G. Chen, et al., "Traditional Chinese Medicine Compound (Tongxinluo) and Clinical Outcomes of Patients With Acute Myocardial Infarction: The CTS-AMI Randomized Clinical Trial," *Jama* 330, no. 16 (2023): 1534–1545.

80. S. Liu, Z. Wang, Y. Su, et al., "A Neuroanatomical Basis for Electroacupuncture to Drive the Vagal–Adrenal Axis," *Nature* 598, no. 7882 (2021): 641–645.

81. T. H. Kung, M. Cheatham, A. Medenilla, et al., "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models," *PLOS Digital Health* 2 (2022).

82. H. Wang, W. Wu, Z. Dou, L. He, and L. Yang, "Performance and Exploration of ChatGPT in Medical Examination, Records and Education in Chinese: Pave the Way for Medical AI," *International Journal of Medical Informatics* 177 (2023): 105173.

83. M. Reid, N. Savinov, D. Teplyashin, et al., "Gemini 1.5: Unlocking Multimodal Understanding Across Millions of Tokens of Context," *ArXiv* (2024). abs/2403.05530.

84. J. T. Chu, M. P. Wang, C. Shen, K. Viswanath, T. H. Lam, and S. S.-C. Chan, "How, When and Why People Seek Health Information Online: Qualitative Study in Hong Kong," *Interactive Journal of Medical Research* 6 (2017).

85. S. Montagna, S. Ferretti, L. C. Klopfenstein, A. Florio, and M. F Pengo, "Data Decentralisation of LLM-Based Chatbot Systems in Chronic Disease Self-Management," In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good* (Association for Computing Machinery, 2023): 205–212, https://doi.org/10.1145/3582515.3609536.

86. A. Mahmood, S. Cao, M. Stiber, V. N. Antony, and C.-M. Huang, "Voice Assistants for Health Self-Management: Designing for and With Older Adults," *ArXiv* (2024). abs/2409.15488.

87. Z. Ye, C. Tian, Y. Yan, et al., "Implementation Evaluation of Clinical Practice Guidelines for Integrative Medicine," *Medical Journal of Peking Union Medical College Hospital* 15, no. 2 (2024): 413–421.

88. Y. Hong, H. Xu, Y. Liu, et al., "DDID: A Comprehensive Resource for Visualization and Analysis of Diet-Drug Interactions," *Brief Bioinform* 25, no. 3 (2024): bbae212.

89. S. Gerke, T. Minssen, and G. Cohen, "Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare," *Artificial Intelligence in Healthcare* (2020): 295–336.

90. J. Jiao, S. Afroogh, Y. Xu, and C. Phillips, "Navigating LLM Ethics: Advancements, Challenges, and Future Directions," *ArXiv* (2024). abs/2406.18841.

91. J. Pawelczyk, M. Kraus, L. Eckl, et al., "Attitude of Aspiring Orthopaedic Surgeons towards Artificial Intelligence: A Multinational Cross-Sectional Survey Study," *Archives of Orthopaedic and Trauma Surgery* 144, no. 8 (2024): 3541–3552.

92. T. Wu, L. Luo, Y.-F. Li, S. Pan, T.-T. Vu, and G. Haffari, "Continual Learning for Large Language Models: A Survey," *ArXiv* (2024). abs/2402.01364.

93. M. R. Rezaei, R. S. Fard, J. Parker, R. G. Krishnan, and M. Lankarany, eds., Adaptive Knowledge Graphs Enhance Medical Question Answering: Bridging the Gap Between LLMs and Evolving Medical Knowledge (2025), https://arxiv.org/abs/2502.13010.

94. K. Lu, Z. Liang, D. Pan, et al., "Med-R2: Crafting Trustworthy LLM Physicians Through Retrieval and Reasoning of Evidence-Based Medicine," *ArXiv* (2025). abs/2501.11885.

95. L. Luo, J. Ju, B. Xiong, Y.-F. Li, G. Haffari, and P. S. ChatRule, "Mining Logical Rules With Large Language Models for Knowledge Graph Reasoning," *ArXiv* (2023). abs/2309.01538.

96. J. Yang, H. Xu, S. Mirzoyan, et al., "Poisoning Medical Knowledge Using Large Language Models," *Nature Machine Intelligence* 6, no. 10 (2024): 1156–1168.

97. A. Belyaeva, J. Cosentino, F. Hormozdiari, K. Eswaran, S. Shetty, G. C. Corrado, et al., eds. Multimodal LLMs for Health Grounded in Individual-Specific Data (ML4MHD, 2023).

98. S. Xu, L. Yang, C. J. Kelly, et al., "ELIXR: Towards a General Purpose X-ray Artificial Intelligence System Through Alignment of Large Language Models and Radiology Vision Encoders," *ArXiv* (2023). abs/2308.01317.

99. J. Yan, J. Cai, Z. Xu, et al., "Tongue Crack Recognition Using Segmentation Based Deep Learning," *Scientific Reports* 13 (2023): 511.

100. J. Lim, J. Li, X. Feng, et al., "Machine Learning Classification of Polycystic Ovary Syndrome Based on Radial Pulse Wave Analysis," *BMC Complementary Medicine and Therapies* 23, no. 1 (2023): 409.

101. W. C. Lam, A. Lyu, and Z. Bian, "ICD-11: Impact on Traditional Chinese Medicine and World Healthcare Systems," *Pharmaceutical Medicine* 33, no. 5 (2019): 373–377.

102. Q. Xu, "WHO International Standard Terminologies on Traditional Chinese Medicine: Use in Context, Creatively," *Integrative Medicine in Nephrology and Andrology* 10 (2023), https://doi.org/10.1007/s11655-011-0818-8.

103. J. X. Wang, R. Sun, D. X. Si, et al., "Clinical Practice Guidelines of Chinese Patent Medicine in China: A Critical Review," *Complementary Therapies in Medicine* 85 (2024): 103077.

104. J. Zhao, Q. Zha, M. Jiang, H. Cao, and A. Lu, "Expert Consensus on the Treatment of Rheumatoid Arthritis With Chinese Patent Medicines," *The Journal of Alternative and Complementary Medicine* 19, no. 2 (2012): 111–118.

105. Q. Lv, G. Chen, H. He, et al., "TCMBank-the Largest TCM Database Provides Deep Learning-Based Chinese-Western Medicine Exclusion Prediction," *Signal Transduction and Targeted Therapy* 8, no. 1 (2023): 127.

106. Y. Chang, P. Kang, T. Cui, et al., "Pharmacological Inhibition of Demethylzeylasteral on JAK-STAT Signaling Ameliorates Vitiligo," *Journal of Translational Medicine* 21, no. 1 (2023): 434.

107. F. Cheng, X. Wang, W. Song, et al., "Biologic Basis of TCM Syndromes and the Standardization of Syndrome Classification," *Journal of Traditional Chinese Medical Sciences* 1, no. 2 (2014): 92–97.

108. H. H. Zhao, J. X. Chen, and Q. Shi, "Gel Electrophoresis Analysis on Plasma Differential Protein in Patients With Unstable Angina of Blood-Stasis Pattern," *Zhongguo Zhong Xi Yi Jie He Za Zhi Zhongguo Zhongxiyi Jiehe Zazhi* 30, no. 5 (2010): 488–492.

109. Q. Qiu, C. Li, Y. Wang, et al., "Plasma Metabonomics Study on Chinese Medicine Syndrome Evolution of Heart Failure Rats Caused by LAD Ligation," *BMC Complementary and Alternative Medicine* 14, no. 1 (2014): 232.

110. Y. Gu, C. Lu, Q. Zha, et al., "Plasma Metabonomics Study of Rheumatoid Arthritis and Its Chinese Medicine Subtypes by Using Liquid Chromatography and Gas Chromatography Coupled With Mass Spectrometry," *Molecular Biosystems* 8, no. 5 (2012): 1535–1543.

111. F. Bahari and M. Yavari, "Hot and Cold Theory: Evidence in Systems Biology," in *Hot and Cold Theory: The Path towards Personalized Medicine*, ed. M. Yavari (Cham: Springer International Publishing, 2021): 135–160.

112. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, ed. H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates Inc., 2020): 9459–9474.

113. Ma X, Gong Y, He P, Zhao H, Duan N, eds., *Query Rewriting in Retrieval-Augmented Large Language Models2023 December* (Singapore: Association for Computational Linguistics).

114. D. Arora, A. Kini, S. R. Chowdhury, N. Natarajan, G. Sinha, and A. Sharma, "GAR-Meets-RAG Paradigm for Zero-Shot Information Retrieval," *ArXiv* (2023). abs/2310.20158.

115. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive Mixtures of Local Experts," *Neural Computation* 3 (1991): 79–87.

116. N. M. Shazeer, A. Mirhoseini, K. Maziarz, et al., "Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer," *ArXiv* (2017). abs/1701.06538.

117. Z. Duan and B. Liu, "Promoting the Development of Integrated Traditional Chinese Medicine and Western Medicine With Confidence in Traditional Chinese Medicine," *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* 33, no. 10 (2021): 1153–1156.

118. X. Wang, J. Zhao, E. Marostica, W. Yuan, J. Jin, J. Zhang, et al., "A Pathology Foundation Model for Cancer Diagnosis and Prognosis Prediction," *Nature* 634, no. 8035 (2024): 970–978.

119. H. Xu, N. Usuyama, J. Bagga, et al., "A Whole-Slide Foundation Model for Digital Pathology From Real-World Data," *Nature* 630, no. 8015 (2024): 181–188.

**Supporting Information**

Additional supporting information can be found online in the Supporting Information section.