When Traditional Medicine Meets AI: Critical Considerations for AI-Empowered Clinical Support in Traditional Medicine

YULING SUN*, Fudan University, China
WENJING YUE*, East China Normal University, China
XIAOFU JIN, Hong Kong University of Science and Technology, China
SHUAI MA, Aalto University, Finland
XIAOJUAN MA, Hong Kong University of Science and Technology, Hong Kong SAR
XIAOLING WANG†, East China Normal University, China

Traditional Medicine (TM) is the oldest healthcare form and has been increasingly adopted as the primary or complementary medical therapy in the world. However, TM's practical development remains highly challenging. While artificial intelligence (AI) has become powerful in advancing modern medicine, limited attention has been paid to its potential and usage in TM. This study addresses this gap through a probe-based interview study with 16 TM clinicians, examining their experiences, perceptions, and expectations of AI-empowered clinical support systems. Our findings reveal that despite numerous AI-CDS systems, their practical usage in TM settings was still limited. We identify a series of practical challenges when integrating AI-CDS into TM clinical scenarios, largely due to TM's unique features and the significant data work challenges these features present. We end by critically discussing the potential issues that may arise when integrating AI into practical TM scenarios, and proposing a series of practical recommendations for future studies.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**;

Additional Key Words and Phrases: Clinical Decision-making, Traditional Medicine, Traditional Chinese Medicine, Artificial Intelligence, Clinician, Adoption, Usability, Qualitative

ACM Reference Format:
Yuling Sun, Wenjing Yue, Xiaofu Jin, Shuai Ma, Xiaojuan Ma, and Xiaoling Wang. 2025. When Traditional Medicine Meets AI: Critical Considerations for AI-Empowered Clinical Support in Traditional Medicine. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW524 (November 2025), 37 pages. https://doi.org/10.1145/3757705

1 Introduction

Traditional Medicine (TM), also known as indigenous medicine [117], is defined by the World Health Organization (WHO) as "*the sum total of the knowledge, skills, and practices based on the theories, beliefs, and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health as well as in the prevention, diagnosis, improvement, or treatment of physical and*

*The first two authors contributed equally to this work.
†Corresponding author

Authors' Contact Information: Yuling Sun, Fudan University, China, yulingsun@fudan.edu.cn; Wenjing Yue, East China Normal University, China, wjyue@stu.ecnu.edu.cn; Xiaofu Jin, Hong Kong University of Science and Technology, China, xjinao@connect.ust.hk; Shuai Ma, Aalto University, Finland, shuai.ma@aalto.fi; Xiaojuan Ma, Hong Kong University of Science and Technology, Hong Kong SAR, mxj@cse.ust.hk; Xiaoling Wang, East China Normal University, China, xlwang@cs.ecnu.edu.cn.

ACM 2573-0142/2025/11-ARTCSCW524
https://doi.org/10.1145/3757705

mental illness" [73]. Compared to more standardized, universal, and specific disease-oriented modern medicine [35, 44], TM is often experience-based, personalized, and holistic [75, 137], developed over generations within the folk beliefs of various societies, including indigenous peoples, before the era of modern medicine [27]. Research has provided strong evidence that TM holds significant effectiveness for health improvement as well as the prevention and treatment of various diseases [15, 42], especially complex, internal and chronic diseases such as autoimmune disorders, cardiovascular diseases [15, 68] and cancers [84, 142], diabetes [3, 14], hypertension [31, 37], pain [15], etc. Given these benefits, a growing number of people adopt TM as the primary or complementary medical therapy. According to WHO's report [9], in many Asian countries, especially developing countries, such as China [102], India [107], Kenya [11, 90], Ethiopia [109], more than 80% of the population rely on TM to meet their primary health care needs [47]. In developed Asian countries, such as Japan, South Korea and Singapore, 65% of people also use TM in primary health care [9]. Elsewhere around the world beyond Asia, such as the United States, Australia, and some European countries, TM has also been progressively adopted as the significant Complementary and Alternative Medical therapy (CAM) [15, 72].

As an increasingly popular medicine and valuable cultural heritage in the world [15, 68], the development of TM has received increasing attention globally. In 2013, WHO launched the "WHO traditional medicine strategy 2014–2023" [73] to promote TM's development and strengthen the role it plays in maintaining human health. Several countries (e.g. China [124], Japan [131], South Korea [75], and Malaysia [15]) have also issued policies to support the development, application and inheritance of their TM systems. Even so, the practical development of TM remains slow and challenging. Besides the objective socio-cultural and environmental factors (e.g. regulation, plummeting plant resources [28]), TM's unique characteristics such as experience-based treatment methods [137], lack of standardization [91], intricate clinical reasoning and decision process [44], etc., have posed significant challenges to its practical development.

In recent years, AI-empowered technologies have been increasingly integrated into medical settings, reshaping medical diagnostics and treatment paradigms, and promoting the practical development of medicine, particularly in modern medicine. In TM, a growing body of recent research has developed AI-CDS technologies to, for instance, mine knowledge [2, 23], recommend drugs [2, 59], assist diagnosis [12, 13, 129], etc. Various AI-CDS systems and tools (e.g., TCMISS [103], LLM-based Qibo [136] and ShennongMGS [22], and TCM-FTP [140]) have also been developed and deployed to support TM's practice and development.

Despite the seminal studies on the technical aspects of applying AI-CDS to enhance the performance and effectiveness of TM practices, very little is published to examine the adoption and usage of these technologies in practical TM settings. This research gap is critical because, prior literature has provided strong evidence that, AI-CDS only can be effective when they are adopted by practitioners and appropriately integrated into their practical workflow [36, 52]. Researchers therefore have articulated an urgent need to better examine medical AI technologies in the real-world clinical context [58, 111]. Particularly, compared to modern medicine, TM's inherent empirical nature [137], complex diagnosis ways (i.e., look-listen-question-feel) [48, 102], and personalized and ambiguous treatments, make AI-empowered TM more complicated and challenging. Therefore, it is more critical to deeply examine how AI-CDS technologies, which have already played significant roles in advancing modern medicine, have been adopted and used in practical TM settings, and how TM practitioners experience and perceive these technologies.

In CSCW and related fields, although there has been surging interest in examining AI-CDS in practical clinical settings (e.g. [36, 110]), most attention has been paid to modern medicine scenarios. To our knowledge, research has yet to examine AI-CDS technologies in practical TM scenarios. Our study fills this gap through a probe-based interview study with 16 TM clinicians, with specific research questions of 1) *whether and how existing AI-CDS technologies have been successfully integrated and used in practical TM settings*, and 2) *how TM clinicians experience and perceive AI-CDS technologies in TM settings*.

Our findings indicate, despite numerous AI-CDS technological efforts and products for TM, the use of these technologies in practical TM clinical settings remains quite limited. Most of our participants expressed a cautious attitude towards this technological intervention in TM, due to TM's inherent characteristics such as the complex reasoning process, reliance on established theoretical frameworks, challenges of explaining and quantifying reasoning and diagnostic process. Particularly, these characteristics pose significant *data work* [80, 96] challenges for clinicians, which impact the effectiveness and usability of AI-CDS tools, as well as clinicals' adoption in TM settings.

Our study contributes to CSCW community by expanding on previous research that examines AI-CDS technologies from the perspective of clinicians, with a specific focus on Traditional Medicine (TM), a widely practiced yet understudied field. Through a probe-based interview study, we provide an in-depth exploration of TM clinicians' experiences, perceptions, and the data challenges they face in adopting and using AI-CDS in practical clinical settings. Drawn on these findings, we critically discuss how and for what AI-CDS technologies could be better designed to enhance the adoption and utilization in practical TM settings, following detailed recommendations for future studies looking into this research area.

2 Related Work

Our study is situated into the research of AI-CDS technologies in medicine, and clinicians' adoption and experience of these technologies. In this section, we review existing work in these research lines, setting up the background for understanding the contributions of our study.

2.1 AI-Empowered Clinical Support Technologies

AI-empowered technologies and systems have been widely adopted and used in practical clinical settings. A variety of AI-empowered technologies have been developed and deployed, spanning from AI algorithms and models (e.g., drug discovery [7], recommendation [32], toxicity prediction [77], medical imaging recognition [74], etc.), to various software (e.g., AI-empowered clinical assistant [123], disease intervention system [20, 30], patients-oriented health counseling systems [66], public health preservation platform [21], etc.) and hardware tools (e.g., devices for monitoring electrocardiograms (ECG) [51], electromyograms (EMG) [64], sleep enhancing [54], etc.), tailored to specific usage scenarios. These technologies play significant roles in enhancing medical development and reshaping practitioners' clinical practices, such as disease diagnosis and prediction, clinical decision support (CDS), personalized treatment, drug discovery, patient health management and prevention, etc.

The recent advent of Large Language Models (LLMs) further reshapes the medical paradigm in contemporary society, introducing novel ways to enhance healthcare delivery, research, and patient support. As they have become increasingly powerful and are capable of simulating natural interpersonal communications, LLMs have drawn significant attention of AI and HCI scholars. Researchers, on the one hand, have designed various LLMs tailored for the medical scenarios to support health question and answers(Q&A) [5, 135], emotional support [4], assisted clinical decision-making [79] and advice [1], etc. On the other hand, various specific LLMs-based applications and tools are being developed and explored in medical scenarios, such as LLM-based voice assistant for communication between healthcare providers and patients [127], personalized healthcare recommendations [29], medical learning assistants [43], LLM-empowered embedded medical system development [24], etc., to improve clinical efficiency and enhance clinical decision-making.

In CSCW and HCI, given the high-stakes nature of medical scenarios, researchers pay much attention to exploring human-centered design methods to increase AI-CDS's practical effectiveness, usability, and system uptake, such as using human-centered information [49] and data visualization [50] to help reduce clinicians' over-reliance on AI-CDS tools while building trust, using stakeholder-engaged design method to deliver data aligned with clinicians' needs for disease assessment and enable key patient insights to be more accessible and understandable for them to make quicker decision-making [25], and improving system interactivity to make it more ease-to-use in clinical practice [79].

These efforts have significantly advanced the performance and usability of AI-CDS in clinical practice settings. Even though, the complexity of real-world clinical scenarios [111] and the diversity of clinical data [81] still challenge the practical usage of AI-CDS. Given the high-risk nature of medical decisions, scholars foreground that, before the large-scale deployment of AI-CDS, it is crucial to conduct in-depth evaluations within real-world clinical scenarios to examine the practical effectiveness and usability, identifying the challenges associated with integrating AI-CDS into real-world practice and informing the design to better fit AI-CDS into clinical practice. Our study joins this research effort to enrich the qualitative examination of AI-CDS in practical TM settings, which remains understudied in existing literature.

2.2 Clinicians' Adoption and Usage to AI-Empowered Clinical Support Technologies

Prior literature has provided strong evidence that, AI-empowered technologies only can be effective when they are adopted by practitioners and appropriately integrated into their practical workflows [36, 52]. Misalignment between technologies and clinical workflows might increase clinicians' workload, reducing their adoption and using experience [6]. Therefore, CSCW and HCI researchers have paid significant attention to examining AI-CDS technologies in real-world clinical contexts from the perspective of clinicians, understanding how they adopt, use, and experience AI-empowered technologies.

Several challenges clinicians face have been identified. Explainability, for instance, is one major challenge influencing clinicians' adoption and usage to AI-empowered clinical support technologies. Since AI technologies are operated as "black box", it is often difficult for clinicians to trust the results provided by models fully [61]. Researchers note that clinicians often struggle to understand how algorithms generate recommendations, which creates inconsistencies in applying these tools in real-world clinical decision-making scenarios [111]. Jacobs et al. [36] further point out that clinicians prefer explanations based on evidence rather than abstract model features. Moreover, no AI model can guarantee perfect accuracy, making clinicians' trust issues in adopting AI-CDS [126]. Additionally, the interaction ways with AI-CDS tools, such as click- or input-based, and image- or text-based, also significantly influence clinicians' experience [52].

With these challenges, researchers have explored various human-centered technological solutions. For example, many Explainable Artificial Intelligence (XAI) technologies, such as Kovalchuk et al. [45]'s three-stage method which integrates decision-making environments with existing literature, and Xie et al. [123]'s CheXplain which provides tiered explanations to meet the varied explanation needs of clinicians across departments, have been proposed to enhance AI-CDS's transparency, helping clinicians better understand model feedback [60]. Some researchers highlighted the role of stakeholders' engagement for better explanation. Ferdush et al. [26], for instance, designed an AI-CDS framework, enabling patients to inquire about their data and understand the reasoning behind these results. Moreover, researchers have proposed evidence-enhanced approaches to building trust of AI-CDS. Yang et al. [126], for instance, proposed a literature-based explainable system, enabling clinicians to verify AI-generated recommendations and persuading them to accept accurate advice while disregarding incorrect feedback.

In terms of interactivity, user-centered design approach is frequently used to provide on-demand explanations and support clinicians' decision-making as well as stakeholders' collaboration. Jacobs et al. [36], for instance, utilize an iterative design process to create a multi-user AI-CDS that supports technology-mediated collaboration between patients and clinicians. Ser et al. [85] emphasize integrating AI-CDS with Electronic Health Records (EHR) and ensuring compatibility with hospital workflows to improve the system's practicality and adoption. Wang et al. [111] demonstrate that input-based interaction methods in AI-CDS can streamline the diagnostic process, saving clinicians' time. These studies foreground the necessity of user-centered interactive design to integrate AI-CDS into practical clinical settings successfully.

These studies have significantly advanced the acceptance and use of AI-CDS among clinical practitioners. However, most existing investigations have been conducted within the context of modern medicine, with limited attention given to TM. It remains unclear how existing AI-CDS technologies function in the context of TM and how TM clinicians adopt and utilize them. Our research fills this gap through a qualitative study with TM clinicians, examining their adoption and experiences with AI-CDS technologies in practical TM scenarios.

2.3 AI-Empowered Clinical Support Technologies in Traditional Medicine

With the high popularity of TM, a large body of research has paid attention to AI-based technology design and development to empower TM's development and applications. Ongoing efforts span from algorithmic exploration (e.g. machine learning and deep learning based TM knowledge analysis and extraction [87, 120], core symptom examining [119], medication patterns recognition [63], drug and prescription recommendation [39, 40, 55], etc.), to image and signal recognition technology based TM assist diagnosis (e.g., tongue diagnosis [62], facial diagnosis [57], olfactory diagnosis [53], and auscultation [125], etc.)

The recent advanced LLMs have also drawn the attention of this research area. Increasing LLM-empowered applications are now being explored and designed to support practical TM, especially in Traditional Chinese Medicine (TCM). Yue et al. [132], for instance, proposed an LLM evaluation benchmark designed explicitly for TCM, highlighting the need for further refinement of general-purpose LLMs to serve this area better. To enhance LLM performance in TM, some researchers fine-tuned general-purposed LLMs with TM literature and knowledge graphs, focusing on tasks such as assisted diagnosis [101, 136], epidemic prevention and treatment [141], as well as the generation of medication usage instructions [56]. Meanwhile, LLM-based systems have also been employed to support TM practices, such as formula classification [115], medication recommendations [78, 140], medication guidance and adverse drug reaction predictions [22].

However, most existing attention has been paid to the technical aspects of AI-CDS technologies for TM. To our knowledge, no attention has been paid to examining the adoption and usage of these technologies in practical TM clinical settings. Given TM's inherent empirical nature [137], complex diagnosis ways [48, 102], and personalized treatments, it is therefore critical to deeply examine how these AI-CDS technologies have been adopted and used in practical TM settings, and how TM practitioners experience and perceive these technologies. Our study contributes to this research gap through a probe-based interview study with TM clinicians.

3 The Study

With the aim of gaining insights into TM clinicians' experiences and perceptions of AI-empowered clinical support technologies in their practical clinical scenarios, we conducted a probe-based interview study [34] by using the technology probe named *Shuzhiqihuang* (数智岐黄, *Shuzhi*

Qihuang)[1]. We now elaborate on our study context, technology probe we used, as well as data collecting and analyzing methods.

3.1 Study Context: Traditional Chinese Medicine

Our study was primarily conducted in China, with a specific focus on *Traditional Chinese Medicine (TCM)*, one of the oldest and most representative forms of traditional medicine globally [15]. TCM has developed its comprehensive theoretical system for diagnosing and treating diseases over thousands of years, employing the diagnostic method known as "look-listen-question-feel" [48] and treatment methods such as acupuncture, herbal medicine, and massage [137]. TCM is also the foundation of many TM systems across Asia [75]. For instance, the traditional medicine practices of South Korea and Japan (referred to as "Kampo" in Japan) can be traced back to ancient TCM texts "Treatise on Cold Pathogenic and Miscellaneous Diseases" and "Canon of Internal Medicine" [131]. Similarly, Vietnamese TM is developed by integrating TCM into their own culture [108], and India's TM (Ayurveda) shares several commonalities with TCM [76].

Currently, TCM is the most widely accepted and influential form of TM worldwide. According to WHO survey [15], TCM has been extensively adopted in and beyond China. In China, according to Chinese TCM healthcare institutions, 1.28 billion people (more than 90% of China's total population) across China in 2023 have adopted TCM diagnoses and treatment [67]. In Japan, South Korea, and Singapore, 60% to 75% of the population consult TCM practitioners annually, while in the U.S., U.K., and Australia, 40%, 10%, and two-thirds of the population, respectively, adopted TCM as the significant complementary and alternative medical therapy [15, 72].

Given this widespread adoption, the development of TCM has drawn significant attention from international organizations (e.g., WHO [73]) and the Chinese government [70, 106]. In China, the development, inheritance, and education of TCM have long been priorities of the government [105]. Particularly in China's vigorous promotion of national digital transformation [93], there has been a strong focus on leveraging AI-empowered technologies to promote TCM. Various AI-empowered technologies (such as crucial diagnosis information mining [63, 119], drug recommendation [59, 140], medical Q&A [136], and medication guidance [22]) have been designed, deployed, and widely implemented for TCM. This creates an ideal research context for our study to examine the practical effectiveness of AI-CDS in TCM.

3.2 Technology Probe: *Shuzhiqihuang*

Shuzhiqihuang is a public AI-CDS system designed for TCM, offering functions of explainable prescription recommendations and LLM-empowered TCM Q&A. It was collaboratively designed by our research teams, including AI and HCI researchers, as well as TCM clinicians and researchers at one TCM University in China. The initial goal of our research team was to design an AI-CDS probe to elicit reflections and discussions from clinicians. To this end, the team collaboratively designed and developed AI-assisted system *Shuzhiqihuang*. Specifically, clinicians contributed their diagnostic and reasoning processes and support data annotation; AI engineers developed algorithms based on expert knowledge; and HCI researchers designed interactive interfaces grounded in the clinical scenarios. Through several rounds of iterative refinement, we finalized *Shuzhiqihuang* as illustrated in Fig. 1, with three core functional modules.

- **TCM Knowledge Visualization (Fig. 2)**. Under the guidance of TCM experts, we constructed a medicine knowledge graph (KG) centered on herbs and symptoms, integrating

[1]In Chinese, 数 (*shu*) refers to data, 智 (*zhi*) refers to intelligence (especially AI), and 岐黄 (*qihuang*) is a classical synonym for traditional Chinese medicine. The term 数智岐黄 generally refers to the integration of data-driven and AI-empowered technologies into TCM practices.

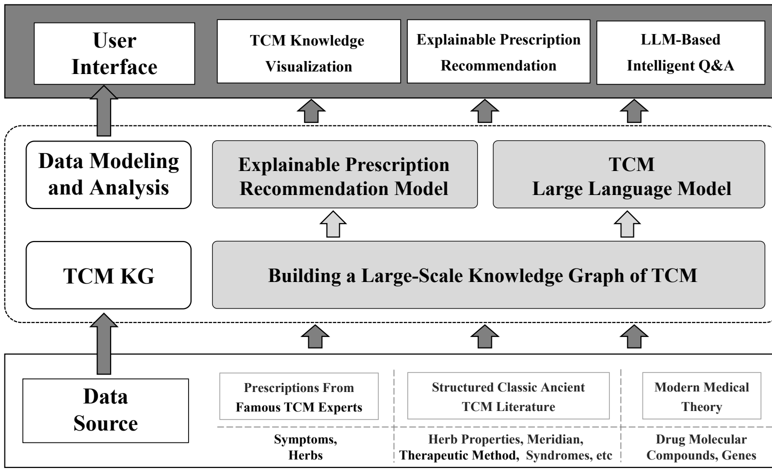


Fig. 1. The system structure of *Shuzhiqihuang*

TCM and Western medicine concepts. The data source of this KG included two structured TCM knowledge databases, semi-structured data from herbal websites, and 26,360 prescriptions from famous TCM experts. Ultimately, the TCM and Western medicine KG contained 123,358 triples, 37,114 TCM entities, and 16 types of relationships.

- **Clinical-AI Collaborative Explainable Prescription Recommendation (Fig. 3).** According to "syndrome differentiation and treatment" and "pharmacological molecular" theory of herbs, we divided TCM decision-making process into three meta-paths, and combined graph neural networks and attention mechanisms to recommend herbs. Meanwhile, we designed a clinical-AI collaborative interface for clinicians to adjust the recommendation results, as well as four-stage clinical diagnostic thought process (Fig. 4) to assist clinicians in understanding the recommendation results.
- **LLM-Based Intelligent Q&A (Fig. 5).** We constructed a TCM LLM with prompt tuning, chain-of-thought (CoT), and retrieval augmented generation (RAG) technologies with 26 unstructured ancient TCM literature. The LLM was tailored for knowledge Q&A in TCM domain, supporting natural conversational interactions, RAG-empowered health Q&A, and prescription recommendations.

In terms of algorithm design, the algorithms implemented in *Shuzhiqihuang* were collaboratively developed by clinical experts and AI engineers on our team, to ensure that the reasoning process and interactive experience align closely with practical TM workflows. The technical details were provided in Appendix B, and the evaluation procedures and results were provided in Appendix C and Appendix D. Based on the conducted quantitative evaluations, the technical performance of *Shuzhiqihuang* had achieved state-of-the-art (SOTA) results in TM prescription recommendation and Q&A benchmark tasks. In terms of interaction design, *Shuzhiqihuang* offers an easy-to-use interface with clear navigation, allowing TCM clinicians and stakeholders to quickly familiarize themselves with it. It also offers a structured text-based interactive interface supporting clinician-AI collaborative prescription recommendation (Fig. 3), a meta-path-based explainable interface providing clinicians explainable recommendation results (Fig. 4), and a natural, LLM-based intelligent Q&A interface (Fig. 5) to meet the diverse interaction needs of TCM clinicians.

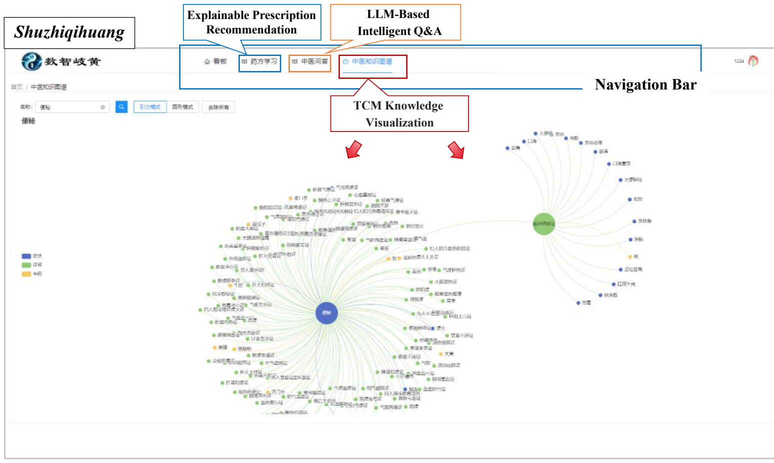


Fig. 2. TCM Knowledge Visualization Interface. Above the system is the navigation bar, with three functions corresponding to three modules. The shown case represents the entities and relationships related to the symptom "constipation".

Fig. 3. Clinicians-AI collaborative explainable prescription recommendation Interface. The use process is divided into five steps: ① Clinicians select data sources, typically ancient literature, from the database. ② They enter the patient's symptoms, with the system's assisted input. ③ After clicking the recommendation button, the system automatically recommends multiple herbs displayed at the bottom of the system, including names, images, and effect descriptions.④ Clinicians click the button on the right side of each herb to view the reasoning path shown in Fig 4. ⑤ If the clinician disagrees with the recommended herbs, they can click the button in the fifth step to remove or adjust the herbs. The remaining herbs will form a prescription stored in the clinician's account prescription library.

To use *Shuzhiqihuang*, users can select the desired function through the navigation bar at the top of Fig. 2. In the prescription recommendation module (Fig. 3), clinicians can directly input the patient's symptoms through the structured text input interface, and the system will automatically

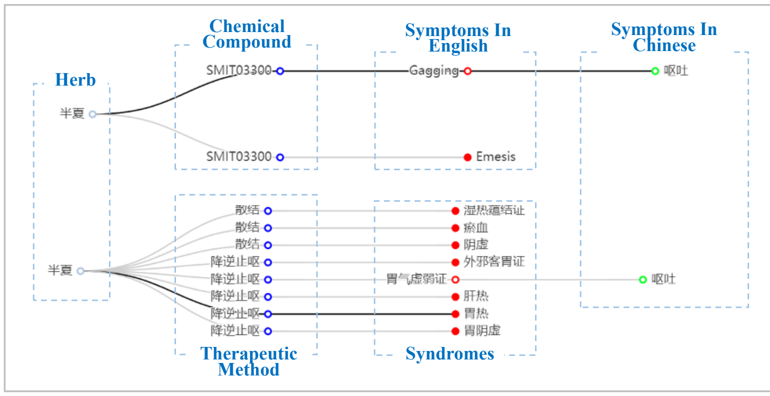


Fig. 4. Explainable reasoning path of herb recommendation. The upper is a reasoning meta-path generated based on pharmacological molecular theory, and the following is a reasoning meta-path generated with treatment based on syndrome differentiation.

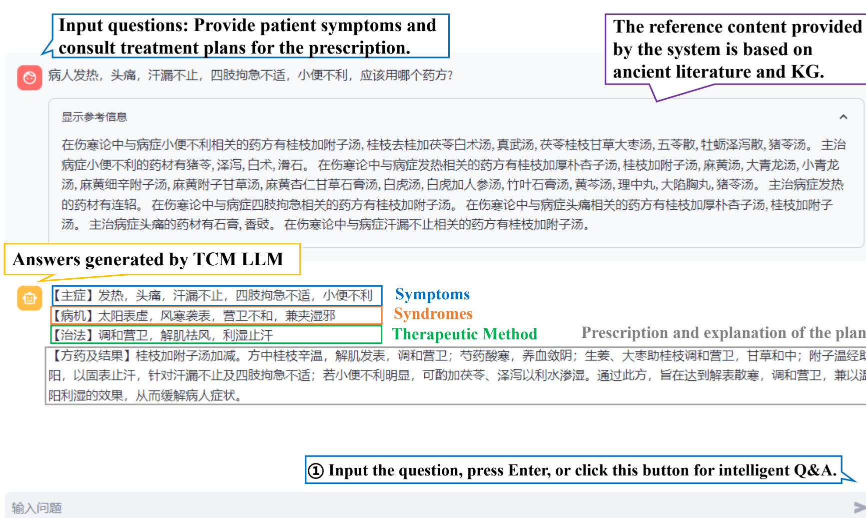


Fig. 5. LLM-based intelligent Q&A Interface

generate recommended prescriptions. Through the clinician-AI collaborative explainable prescription recommendation interface (Fig. 3), clinicians can add, delete, or modify prescription herbs and dosages based on their personal experience. The recommendation results are followed by a meta-path-based explainable path and relevant rationales (Fig. 4). In the LLM-based intelligent Q&A module, users can input natural language questions. The system leverages the LLM to generate answers with corresponding references (Fig. 5). In general, *Shuzhiqihuang* integrates almost all necessary AI-CDS technologies, including knowledge graphs, deep learning, explainable AI, LLMs, and TCM knowledge, to provide clinical decision support tailored for TCM practitioners. We therefore believe *Shuzhiqihuang* an ideal technology probe to investigate AI-CDS technologies for TCM.

Table 1. Basic Information of Participants. *G* in the row headings denotes Gender.

ID	Age	G	Working unit	Department	Position	Edu_background
P1	34	F	Hospital	Dermatology	Attending physician	PhD
P2	38	F	University & Hospital	TCM surgery	Attending physician	PhD
P3	51	M	University & Hospital	TCM internal medicine	Associate chief physician	PhD
P4	28	M	Hospital	Department of acupuncture	Resident physician	Master
P5	33	M	Hospital	Orthopedics and traumatology	Attending physician	Master
P6	47	F	University & Hospital	Clinical outpatient	Lecturer	PhD
P7	28	F	Hospital	Dermatology	Resident physician	PhD candidate
P8	37	F	Hospital	TCM department (gynaecology)	Associate chief physician	Master
P9	31	M	Hospital	TCM surgery	Attending physician	Master
P10	35	F	Hospital	TCM internal medicine	Physician assistant	Master student
P11	39	M	Hospital	TCM internal medicine	Attending physician	Master
P12	37	F	Hospital	Dermatology	Attending physician	Master
P13	58	M	Hospital	TCM internal medicine	Archiater	Master
P14	43	M	Hospital	TCM epidemiology	Archiater	Master
P15	47	M	Hospital	TCM internal medicine	Associate chief physician	PhD
P16	41	F	University	TCM innovation department	Associate researcher	PhD

3.3 Methods

Taking *Shuzhiqihuang* as the technology probe, we investigated our research questions through a qualitative study with 16 TM clinicians in China. The first and second authors, two native Mandarin speakers, collected and analyzed data together. We now present details on the participants, recruitment process, data collection, and analysis process.

3.3.1 Participants and Recruitment. Our participants included 16 TCM clinicians (8 females and 8 males). They were recruited through social network-based recommendations. The institute the authors affiliated had a long-term collaboration with one TCM university in China. Based on the existing collaborative relationship, we asked our partners to help recommend potential clinicians for the study, and recruited 12 participants in this way. Additionally, the first and second authors leveraged their social networks to reach out to friends who worked in the TCM hospital, and recruited the other 4 participants. The basic information of 16 participants is shown in Table 1. Their ages ranged from 28 to 58. 12 worked in the hospital, 1 worked at the TCM university, and 3 worked in both hospital and TCM University. They belonged to various TCM departments, holding diverse professional qualifications and educational backgrounds.

3.3.2 Data Collection. Our data collection process was divided into two main stages, with details shown in Appendix A. In the first stage, we primarily focused on clinicians' current AI technology usage and their experiences with these technologies. In the second stage, we took *Shuzhiqihuang* as a probe to delve deeper into their experiences and perceptions of AI-CDS technologies.

We began by introducing ourselves and the intention of doing research about AI-CDS for TM. After obtaining participants' consent, we conducted semi-structured interviews for the first stage, asking their general work content, AI-empowered technology usage, and experiences. If participants had used one or more AI-empowered technologies, we asked detailed questions about their adoption and usage processes, scenarios, experiences, encountered challenges, and expectations. For those who had not used AI technologies, we asked about their attitudes, perceptions, and expectations to them. In the second stage, we introduced the background, motivation and functionalities of *Shuzhiqihuang* to participants and invited them to explore and evaluate each module firsthand. We specifically invited them to test *Shuzhiqihuang*'s human-AI collaborative prescription recommendation and intelligent Q&A features, through inputting patient symptoms and evaluating the system-generated recommendations from a professional perspective, particularly in terms of accuracy, professionalism, usefulness, trustworthiness, and willingness to use. We pre-established six sets of patient symptoms around common health issues (see more details in Appendix A), and participants could also randomly pose questions based on their expertise. In initial study design, we included five-point Likert-scale questions to quantify participants' evaluations. However, after beginning data collection, most participants expressed that it was difficult to provide quantitative ratings, with the underlying reasons discussed in Section 4.1.1. As a result, we discontinued the collection of quantitative scale data. Instead, after participants fully experienced the functionalities of *Shuzhiqihuang*, we conducted semi-structured interviews to gather their in-depth experiences and perceptions of *Shuzhiqihuang* as well as their broader views on AI-CDS applications in TM.

The interview included questions about their experiences with the quality of *Shuzhiqihuang*'s recommendation and Q&A results, the interface and interaction, the experienced benefits and inconveniences, the possible application scenarios, etc. During the interviews, if participants shared any interesting points or prior experiences, we probed for more details and concrete examples. All questions were designed to be general to ensure inclusivity and encourage participants to share various levels of experiences and perceptions. 7 participants were interviewed through Tencent Meeting and 9 were in-person. Each interview lasted for 40-120 minutes, with 500 RMB compensation. With participants' permission, all interviews were audio-recorded and later transcribed in Chinese verbatim for analysis.

3.3.3 Data Analysis. We followed an inductive thematic analysis to analyze our data [16]. We first transcribed the interview recordings in Chinese. Based on these transcriptions, two authors began independently open coding the data by closely reading all the data and collaboratively developing an initial set of codes. This process was iterated upon by generating an initial codebook from the collected data, classifying different codes into similar themes, discussed these codes and themes on a weekly basis to ensure reliability, and checking and elaborating these codes and themes as new data were obtained. The data collection process stopped once all the core variables reached saturation. By the end of open coding phase, we generated an initial set of codes, capturing clinicians' experiences and evaluations with *Shuzhiqihuang*, including the quality of recommendation and Q&A results, the interface and interaction, the experienced benefits and inconveniences, the possible application scenarios, the limitation of the functions, and the reasons for the limitation, etc.; their experiences with the other technologies they have used, the experienced benefits and inconveniences; their general attitudes, perceptions and expectations to AI-CDS technologies in TM.

Based on these preliminary themes, we then conducted another round of focused coding to refine the initial codes and trace the connections and relationships among them. We particularly reviewed relevant TCM theory documents and literature regarding AI and TCM, which allowed us to further contextualize the relationships among different themes. By refining the codes and themes in relation to the broader theoretical discourses, all data and concepts gradually converged on four interrelated themes: their usage, experiences and evaluation to *Shuzhiqihuang* and other AI-CDS tools, *Shuzhiqihuang*'s usability challenges due to TM's unique features, *Shuzhiqihuang*'s usability challenges due to the incompetent data work, and clinicians' expected using scenarios of AI-CDS. We now present the full details of our findings, using representative quotes, which were then translated into English.

3.4 Ethical Considerations

Our study is approved by the ethics committee of the first authors' institution. During the study, we took careful steps to protect user privacy and preserve the ethics of research. Before starting our study, we informed participants of our intention and background information, and got permission from them. All data collected during our study was used in an anonymous way, i.e., there was no link between the collected data and an individual user. We strictly adhered to the confidentiality agreement, which promises the data is only used for this research and any form of information leakage is not permitted.

4 Findings

Generally, our findings indicated, despite numerous AI-empowered technological efforts and products for TM, the use of these technologies in practical TM clinical settings remains quite limited. Most of our participants expressed that they didn't use many AI-CDS technologies in their clinical practices (Section 4.1). They expressed a cautious attitude towards *Shuzhiqihuang* and existing AI-CDS in TM, due to TM's unique features (Section 4.2) as well as incompetent data work (Section 4.3). Additionally, our participants highlighted other expectations for AI-empowered technologies beyond decision-making (Section 4.4). We now elaborate on these findings in more detail.

4.1 TM Clinician's Existing Usage, Experience and Perception to AI-CDS

4.1.1 Adoption and Perception to Shuzhiqihuang and AI-CDS in TM. Our study revealed that the practical usage of AI-CDS tools among our participants was quite limited. Most technologies they had used were essential medical informatics tools, such as Hospital Information Systems (HIS) (all), TCM literature analyzing platforms (P1, P2), and TCM inheritance system (e.g. [103]) (P1). Aside from these, our participants had not previously used more advanced AI-CDS systems and tools. Further, they expressed varied attitudes toward *Shuzhiqihuang* for TM, influenced by factors such as clinicians' age and workplace environment. Specifically, younger clinicians (e.g., P1, P5, P7, P9, and P10) showed a more positive attitude to *Shuzhiqihuang* and AI-CDS and desired such tools to better assist their practical work. As P1 noted, "*in clinical practice, we may encounter various problems. If there are advanced tools to help improve our decision-making capabilities and clinical skills, that would be great.*" In contrast, senior clinicians (e.g., P15) were less receptive, emphasizing the importance of their personal experience. As P15 explained: "*Senior TM clinicans, such as deputy chief physicians and chief physicians, have our own experience and methods, which AI may not achieve.*"

Meanwhile, the nature of the institutions where clinicians work (i.e., public or private[2]) also influenced TM clinicians' perception to AI-CDS. Specifically, clinicians from public hospitals and traditional clinics generally expressed neutral (e.g., P4, P8, P11) even opposing (e.g., P15) attitudes towards AI-CDS, with reasons of "*it might jeopardize the medical profession*" (P1), "*who would take responsibility if problems arise?*" (P1), "*increasing the technical burden on clinicians*" (P2), "*how to integrate with existing HIS?*" (P4), "*troublesome interaction*" (P5, P7), and "*reducing patient trust in clinicians*" (P5, P6). On the contrary, clinicians from research-oriented private institutions showed very positive support for developing AI-CDS in TCM. For instance, P14, a senior TCM chief physician at a private TCM institute, expressed a high interest in using AI to record his knowledge, thinking, and diagnostic processes. He hoped such a kind of AI-CDS could "*replace me in doing some basic work, then I can focus on more important tasks, like promoting the development of TCM*

[2]In China, hospitals can generally be divided into public and private ones. Many TCM practitioners establish their own private clinics, with different focuses such as research or providing personalized treatments.

internationally". Currently, he has already begun using AI tools like iFLYTEK[3] to record his teaching and diagnostic processes, hoping to build his personalized AI agent.

The academic nature of TCM also influenced clinicians' perceptions of AI-CDS. For example, TCM dermatologists (P1, P7) and internists (P3, P10) expressed positive expectations for AI-CDS, considering it can help analyze vast amounts of ancient texts and literature. In contrast, clinicians in departments requiring hands-on procedures, such as acupuncture (P4), saw limited potential for AI-CDS, with P4 noting that: "*acupuncture department may develop more slowly because we need to control the needling, which is more difficult (for AI).*"

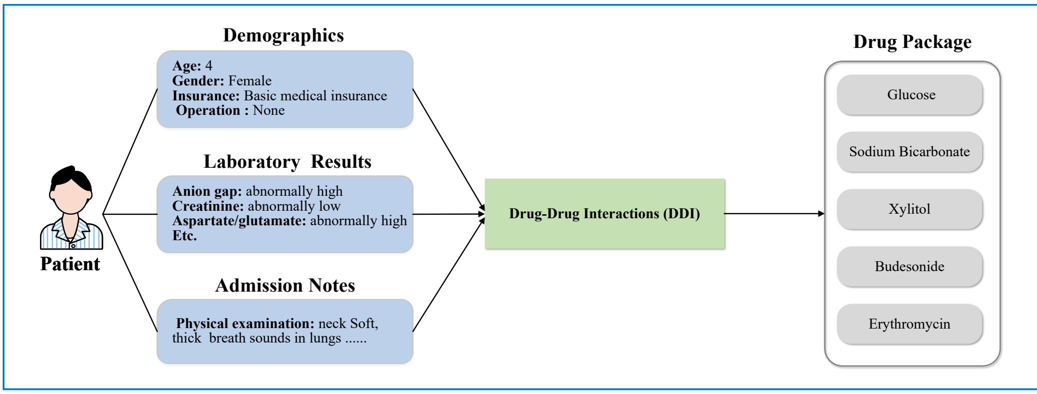
4.1.2 Experience of Shuzhiqihuang. After experiencing and testing *Shuzhiqihuang*, almost all participants expressed it was hard for *Shuzhiqihuang* to support the real-world TM decision-making due to the limitation in accuracy, professionalism, usefulness, and completeness, e.g. "*It's definitely not professional*" (P3), "*This is a textbook answer, true, but of little practical use*" (P6), "*There's no dosage, so it's completely meaningless*" (P16), etc. Some participants noted that they were unable to evaluate the results due to "*lack of key information*" (P4), "*lack of personalized information*" (P6), and "*lack of more detailed information*" (P5). This result was unexpected because, as elaborated in Appendix C and Appendix D, the embedded algorithms of *Shuzhiqihuang* demonstrated state-of-the-art (SOTA) performance on benchmark tasks for TCM prescription recommendation and Q&A. This highlighted that integrating AI technologies into TM clinical scenarios went far beyond algorithmic performance to a range of TCM-specific challenges. Based on insights from our participants, we categorized these challenges into two major types: the usability issues caused by TM's unique features, and data work challenges TM clinicians encountered. Meanwhile, they expressed strong interest in using *Shuzhiqihuang* for other scenarios, such as TM inheritance and education, clinical work assistance, and TM knowledge innovation. We now turn to discuss these challenges and expectations our participants reported.

4.2 TM's Unique Features Challenge the Usability of AI-CDS

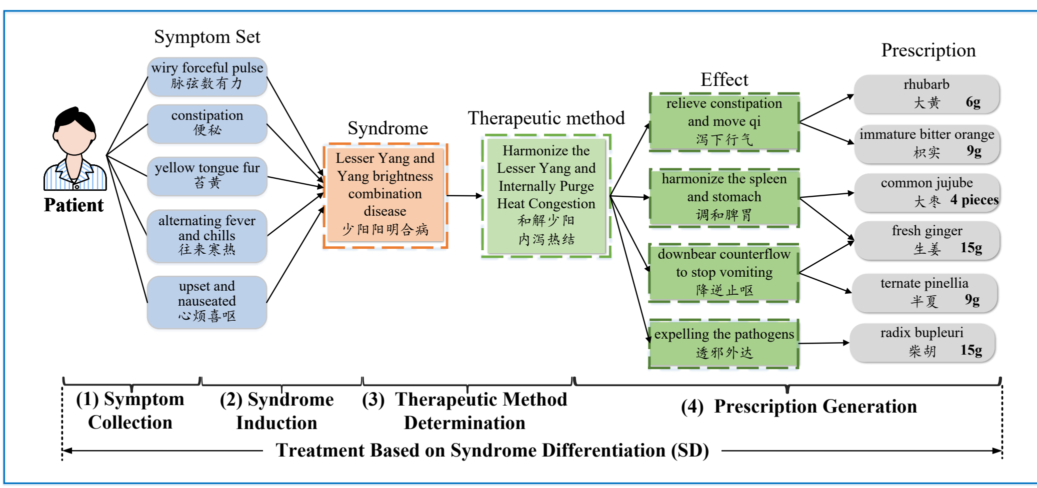
Our participants believed that TM and modern medicine represented two distinct medical systems. TM has unique theoretical foundations, reasoning and diagnostic processes. However, current AI-CDS tools were primarily designed around modern medicine's thinking and reasoning process, causing significant usability challenges when applied in TM scenarios. In this section, we will elaborate on the unique features of TM reported by our participants, which brought the usability challenges of AI-CDS in TM field. Among these challenges, some could be potentially addressed through ongoing technological efforts, such as multi-stage complex reasoning in Section 4.2.1 and socio-contextual embedded comprehensive reasoning in Section 4.2.2, while some others, such as existing theory-based diagnostic in Section 4.2.3 and experience-based diagnostic in Section 4.2.4, were more difficult to resolve, requiring more customized technological designs.

4.2.1 Multi-Stage Complex Reasoning. The reasoning and diagnostic process in modern medicine was relatively linear, explicit, and standardized, as illustrated in Fig. 6a. Clinicians typically began by collecting the patient's demographic and sign information using standardized medical terminology. Then they performed standard inspection processes to obtain definitive laboratory results, informing the prescription of medications based on these results and established clinical guidelines. In Fig. 6a, for instance, the results showed that the patient's anion gap exceeded the standard range (8 $\sim$ 16 mmol/L), then identified as abnormally high. Based on this result, the diagnosis and the corresponding drug were deduced from the patient's demographic information, as well as drug-drug interactions (DDI) considerations.

[3]iFLYTEK (https://peiyin.xunfei.cn/) is a well-known intelligent voice service tool in the Asia Pacific region



(a) Clinical decision process in modern medicine



(b) Clinical decision process in TCM

Fig. 6. The example of clinical decision process in modern medicine (upper) and TCM (lower). The dashed box indicates the absence of standard information like syndromes and therapeutic methods in medical records.

Compared to modern medicine, our participants emphasized that the diagnostic logic of TM was a non-linear, complex, and multi-staged process, known as "syndrome differentiation and treatment (*bian zheng lun zhi*)". P2 described this process as "*first to determine the symptoms through look-listen-question-feel, then to induce the syndrome, followed by determining the therapeutic method and prescribing the herbs.*" According to our participants' description, we illustrated TCM clinical decision process as Fig. 6b. Specifically, the process included:

- **Step 1. Symptom Collection.** TCM practitioners determined patients' symptoms through look-listen-question-feel.
- **Step 2. Syndrome Induction.** The clinician inferred a syndrome based on all symptoms. This was a comprehensive assessment of a symptom set. In Fig. 6b, the clinician identified the syndrome with five symptoms as "Lesser Yang and Yang Brightness Combination Disease".

- **Step 3. Therapeutic Determination.** The clinician determined the corresponding therapeutic based on the syndrome. In the example of Fig. 6b, the physician chose the therapeutic "Harmonize the Lesser Yang and Internally Purge Heat Congestion" to treat the syndrome.
- **Step 4. Prescription Generation.** Finally, the prescription was formulated according to therapeutic, including selecting appropriate herbs and the dosages of each herb.

Further, even each step involved multi-dimensional and many-to-many relationships. For example, during "syndrome induction" step, clinicians determined whether the symptom set was excess or deficiency. As P6 explained: "*these diseases are all vomiting and diarrhea, which some are excess syndromes, while others are deficiency. When multiple syndromes appear simultaneously, we must be clearly distinguished as primary, secondary, or concurrent.*" In addition, "*treating the same disease with different methods*" and "*treating the different disease with the same methods*" often occurred in TM diagnostic process.

In the reasoning process from therapeutic to prescription generation, our participants described, the therapeutic did not directly translate to specific herbs but to therapeutic effects. A single therapeutic often led to multiple effects, and achieving these effects often required a combination of various herbs. Consequently, a therapeutic represented the overall therapeutic outcome of multiple herbs, creating a non-linear relationship between the method and the herbs. Moreover, the herb itself was also complex in terms of components and dosages. As P2 noted, "*Western medicine compounds are relatively straightforward, but in TM, each herb contains numerous components. It is not just a single chemical component but many that work together. The dosage of each herb can also alter its effects.*" Compared to modern medicine's relatively linear sign-to-result reasoning, TM's complex, many-to-many relationship posed significant technological challenges for AI-CDS tools.

4.2.2 Socio-Contextual Embedded Comprehensive Reasoning. Meanwhile, our participants emphasized TM's reasoning process was socio-contextual information-embedded comprehensive process, meaning that TCM's diagnosis depended not only on quantifiable vital signs or symptoms but also on the broader socio-contextual information, such as patient's geographical location, climate, and social environment, i.e., "treatment according to the season, region, and individual's condition" (P1). P2 provided a weather-related diagnostic example to illustrate this social-context embedded reasoning process.

> "*... if it is raining today and you have a headache, the diagnostic decision might differ from that for a headache on a sunny day. The reason for a specific classic prescription might be 'that year had a lot of rain and dampness'. Shanghai is also very damp, so those prescriptions can remain relevant.* "(P2)

However, this social and environmental information was often absent in current data records, leading to AI technologies that fail to align with the practical TCM clinical reasoning process. When P14 queried *Shuzhiqihuang* with "*what to do when skin itching and fever*", he criticized the response and suggested "*to add the crucial information like region (Shanghai) and season, which is the basic of TCM reasoning*". As a TCM epidemiology expert, P14 thus suggested integrating multi-dimensional social-contextual information into the reasoning process of AI models, including "*time factors, seasonal factors, solar terms, meteorological data, satellite cloud images, sensory data, air conditioning usage, spatial factors, urban living conditions, and interpersonal environmental factors.*"

4.2.3 Well-established Theory-Driven Diagnostic. Another key distinction between TM and modern medicine reported by our participants lied in their diagnostic foundations. That was, TM diagnostic relied heavily on well-established theoretical systems, which had been evolved and refined over thousands of years. Compared to modern medicine's rapidly knowledge evolving, our participants described TM's theoretical knowledge as "*a continuous lineage passed down from ancient times* "

(P3). They highlighted the significance of these established theoretical systems for TM reasoning, such as the four great classics of "*'Treatise on Febrile Diseases', 'Essential Prescriptions of the Golden Coffer', 'Canon of Internal Medicine', and 'Detailed Analysis of Epidemic Warm Diseases'*" (P6). Our participants stated that these were "*the crystallization of the wisdom and experience of our ancestors*" (P2) and "*the most critical theoretical foundation of TCM reasoning*" (P6). Despite the socio-environmental changed over time, the patterns and principles summarized by this ancient literature continued to inform TCM clinical practice.

This well-established theory-driven feature of TM significantly influenced the effectiveness and usability of existing AI-CDS models which were often constructed based on modern medicine's reasoning and knowledge developing process, that was, extracting insights and knowledge from large-scale datasets. Yet, our participants emphasized that TM was centered on applying and refining established theories. As P14 explained:

> "*Modern medicine is still building its own knowledge system, as it has yet to understand the relationships between its various elements fully. Its focus is still on how to break boundaries and construct a unified knowledge framework. In contrast, TM already possesses a unified and established knowledge system. AI does not need to construct a new system. Instead, it should ensure its logic aligning with the existing knowledge system. However, current AI-CDS tools in TM often follow the mindset of modern medicine. Then we entered the state of buying computing power and buying data.*" (P14)

Given this perspective, P14 and other participants suggested that AI-CDS tools for TM should first learn the foundational knowledge of TM. Further, this knowledge should be derived from classical, small, high-quality data rather than from extensive datasets of uncertain quality, which might caused "*learning mistakes from mistakes*" (P5) (see Section 4.3.2 for details).

4.2.4 Experience-Based Diagnostic. TM was fundamentally an empirical medical practice [75] where clinicians' personal subjectivity heavily influenced the reasoning and diagnostic results. This subjectivity influenced various aspects of the diagnostic process, such as symptom identification, syndrome induction, and herbs prescribing. Among these, syndrome induction, as illustrated in Figure 6b, was regarded as the most critical step, reflecting the experience-based nature of TM diagnostics. Participants highlighted that it was common for two TM clinicians to prescribe different prescriptions for the same symptom, with "*less than 50% overlap between their prescriptions*" (P1). This divergence occurred due to "*their different paths on syndrome induction*" (P1). As P4 explained, "*One may induct from this direction, while another may induct from another direction. The standards are unclear, leading to varied outcomes.*" Additionally, our participants also pointed out that different clinicians prioritized different diagnostic ways around look-listen-question-feel, and "*some clinicians prioritize 'looking', while others prioritize 'feeling the pulse'*" (P14). These variations were not a matter of right or wrong but reflected "*clinicians' personal preference*" (P1).

In addition to the influence of personal subjective experience, many clinicians we interviewed highlighted the concept of "schools" within TCM. Specifically, TCM had developed various schools and theoretical branches over its long history, each with distinct approaches to diagnosis and treatment. These differences had a significant impact on both diagnosis and herb prescription. P1, for instance, used the diagnostic process of "*damp-heat diseases*" to illustrate how different schools held divergent preferences for prescribing herbs:

> "*Some clinicians from 'Li Dongyuan's Strengthening Spleen School'*[4] *believe that spleen deficiency leads to dampness. Therefore, they prescribe herbs to strengthen the spleen first. However, clinicians from other schools might prioritize reducing dampness directly.*" (P1)

[4]A TCM school, which emphasizes the role of spleen and stomach in healthcare

Fig. 7. The classic ancient literature "Treatise on Febrile Diseases"

Our participants emphasized that whether these differences were brought from clinicians' personal experiences or the development of various schools, there was no definitive right or wrong. Instead, they often "*lead to the same destination through different paths*" (P16). However, such reasoning and diagnostic processes, which relied heavily on differentiated experiences without a clear standard, contrasted sharply with the more objective and standardized reasoning processes of AI-CDS technologies. Our participants considered this as one crucial reason that existing AI-CDS tools such as *Shuzhiqihuang* struggled to be effective in TM.

4.3 Incompetent Data Work Challenges the Usability of AI-CDS Tools for TM

In recent years, TCM had started its informatization and digitalization process, using AI to drive practical advancements [65]. In this process, data was the most critical factor determining the effectiveness of AI technologies. Given the specialized and complex nature of the medicine, much of this data relied heavily on clinicians' extensive data work [80]. However, our study indicated, due to TM's nature of experience-based and complex reasoning processes, TM clinicians often encountered significant challenges in practical data work, which in turn impacted the construction and usability of AI-CDS for TM. We now elaborate on these data work challenges in more detail.

4.3.1 Data Work Challenges in Identifying and Annotating Ancient Literature. As discussed in Section 4.2.3, TCM was built on a comprehensive theoretical system developed over thousands of years. Many ancient TCM texts, such as "Treatise on Febrile Diseases" (Fig. 7), served as the most direct carriers of these classical theories and remained central to the practice and development of today's TCM. Our participants suggested they "*should be preferentially learned by AI technologies*" to ensure an accurate transfer of knowledge. This learning process required careful cleaning and annotation of this ancient work. However, the evolving meanings and dynamic applications of terms and texts over time, influenced by historical and cultural contexts, presented significant challenges for this data work, which, in turn, hindered the effective integration of such knowledge into AI models. As P3 pointed out:

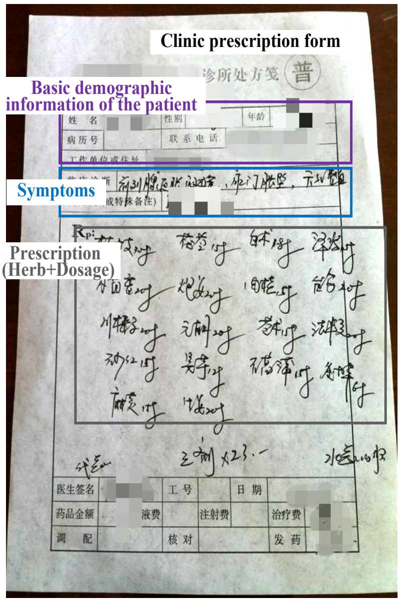
> "*The data in ancient TCM works are primarily textual, not numerical. The writing style and meaning of words in these works can vary across different dynasties, and descriptions of diseases may also differ. We have applied OCR technology to recognize ancient TCM texts, but the recognition accuracy remains relatively low.*" (P3)

Particularly, according to our participants, one essential difference between TM and modern medicine was that modern medicine focused on pathogens and relied heavily on specific quantitative data, making it less sensitive to text information like symptoms. In contrast, TCM was highly responsive to text-based symptoms like headaches and fever. However, as discussed in Section 4.2.4, the evolution of TCM led to the emergence of various theoretical branch and schools, each producing its own literature. While these works shared the common foundational theory, they often used different terminologies for the same symptoms. For example, different ancient texts might describe fever as "*wind-heat*", "*damp-heat*", or "*deficient heat*". This variability posed significant data work challenges, such as data cleaning, annotation, and alignment, requiring extensive manual effort from numerous experts. Our participants indicated that very few doctors were willing to engage in such data work.

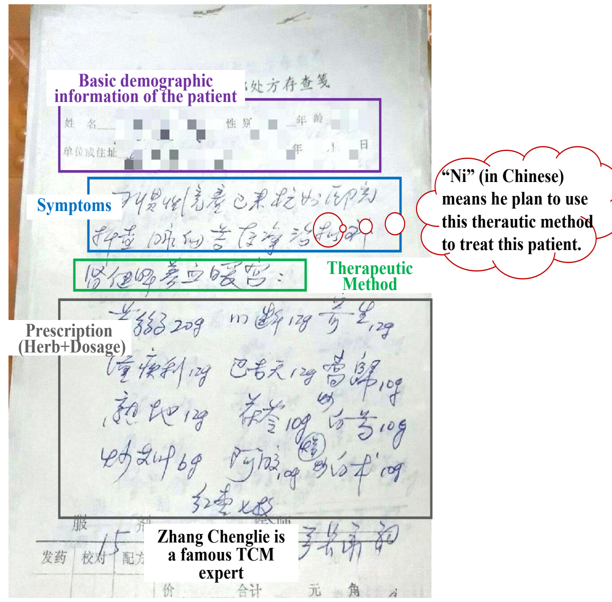
4.3.2 Data Work Challenges in Completely Recording Diagnostic Process. Despite the strong digital infrastructure in contemporary society, constructing high-quality data for TM remained far more complex than for modern medicine. As P16 noted, "*Modern medicine has accumulated incredible data over the years, generating new data daily. All examination results correspond to data, so AI models can continue to be trained. But this process is challenging in TM.*" (P16). Specifically, in TM, on the one hand, much of the symptom data was text-based and needed to be perceived through the clinician's look-listen-question-feeling. This made it difficult to automatically collect data using modern equipment. On the other hand, while there were some digital infrastructure, such as EHR, to assist data collection and recording, these tools were largely designed following the reasoning process of modern medicine, which typically included only symptoms and prescriptions (as shown in Fig.8a). However, as discussed in Section 4.2.1, the reasoning process of TCM was multi-staged, involving symptom collection, syndrome induction, therapeutic method determination, and prescription generation. Critical stages such as syndrome induction and therapeutic method determination were almost not recorded in existing EHRs, leading to incomplete data that failed to capture the entire reasoning process. For instance, Fig. 8a, Fig. 8b and Fig. 8c displayed three types of prescription recordings from clinicians, experts, and ancient literature respectively. From these figures, it was evident that regardless of the sources, diagnostic records often lacked critical syndromes and therapeutic information.

Even for symptoms recorded by EHR systems, our participants noted that existing EHR systems typically lacked a comprehensive database of TM symptoms. To generate high-quality records, clinicians needed to manually input all relevant information, which was challenging given their busy schedules. Our participants told us that, in practice, they often selected symptoms that were available in the system. For symptoms not listed in the system, they rarely documented, or "*only documented primary symptoms, omitting secondary symptoms*" (P13). This process led to poor-quality even erroneous records, which could lead to a cascade of subsequent data applications. Our participants pointed out that relying on these poor-quality even erroneous records for AI modeling and knowledge mining was "*learning errors from errors*" (P5), which failed to create truly useful and effective AI-CDS systems.

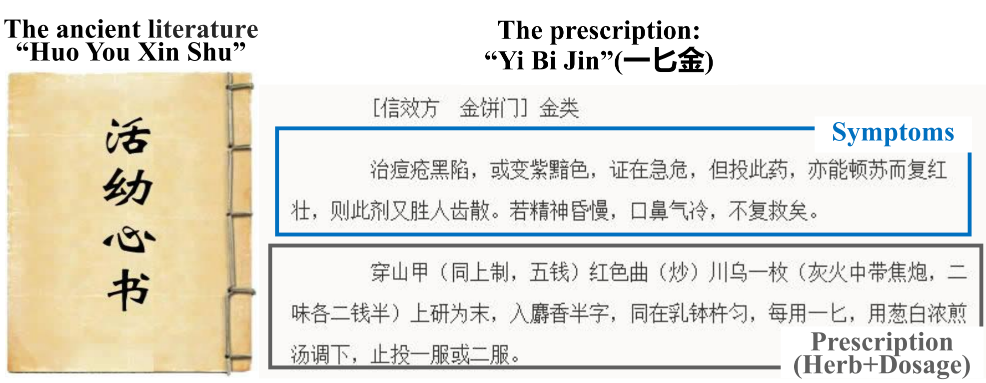
4.3.3 Data Work Challenges in Quantitatively and Standardised Recording Diagnostic Process. The multi-stage reasoning and empirical nature of TCM also presented significant challenges for quantitatively and standardized recording diagnostic processes, which in turn challenged the usability of AI-CDS in TM. Our participants noted that, it was very challenging for them to quantify symptoms and the reasoning process. Specifically, compared to modern medicine, wherein physical signs were often converted into numerical data with defined thresholds, TM data was often qualitative, text-based descriptions, which challenged objective quantification. P2 illustrated this with a concrete example:



(a) The clinic prescription form.



(b) The prescription from a famous TCM expert, Zhang Chenglie.



(c) The classic prescription “Yi Bi Jin” in the ancient literature, “Huo You Xin Shu”.

Fig. 8. Three examples of TCM prescriptions. The black marked area is for additional explanation, the purple marked area is for the patient’s demographics information, the blue marked area is for symptom information, the green marked area is for therapeutic method information (not included in every prescription), and the gray marked area is for prescription information (including information on herb and dosage).

> “*In modern medicine, fever is quantified as a body temperature exceeding by specific temperatures, like 37.2°C or 38.7°C, and a white blood cell count exceeds a threshold indicates an infection. However, TCM describes ‘fever’ as ‘wind-heat,’ ‘damp-heat,’ or ‘deficient heat.’ How do we quantify ‘wind-heat’? Or consider ’stomach qi deficiency,’ what exactly constitutes ‘qi deficiency,’ and how do we measure it?*” (P2)

Moreover, due to the social, regional, and cultural diversity inherent in TCM (as discussed in Section 4.2.2), there was often a lack of mutual recognition among different clinicians from various regions and TCM schools, which further increased the challenges of standardized recording diagnostic. Thus, our participants stressed the importance of standardization at the national level efforts to ensure broader acceptance and trust among practitioners from different backgrounds.

In China, given the state policy support for integrating AI in the inheritance and development of TCM [93], some national standards for TCM had been published, such as "GB/T 16751.2—2021" [94] noted by P5, which encompassed diagnostic criteria, symptoms, and syndromes, along with the WHO published "WHO International Standard Terminologies on Traditional Chinese Medicine"[5]. However, our participants pointed out that the functions of national standards and guidelines for TCM remained relatively limited, particularly in practical application. As P5 explained: "*The current guidelines only introduce what symptoms it has, and what clinical manifestations it presents, but they do not provide other possible texts to match them. We are also struggling with the lack of such matches.*" This gap in practical standards hindered the full potential of AI-CDS tools in TM.

4.4 TM Clinician's Expectation to AI-Empowered Tools

Despite these challenges reported in Section 4.2 and Section 4.3, our participants still highly appreciated the significant value of AI-CDS as *Shuzhiqihuang* we developed. They expressed other expectations for its potential to empower the practical development of TM, which will be detailed in the following.

4.4.1 Learning Partners for Young clinicians. Our participants desired AI tools such as *Shuzhiqihuang* to support guided Q&A, aiding young clinicians in learning TM. Younger participants (such as P4, P5, and P7) noted that they often struggled with the complexity and experiential characteristics of reasoning in TM during the learning process. Particularly, they usually encountered situations where they "*do not understand why their teachers give this prescription*" (P1). They believed that by recreating the reasoning process of clinicians, AI tools such as *Shuzhiqihuang* could significantly aid their learning. As P2 said:

> "*I often followed teachers and copied their prescriptions when I was a student. A common issue was that, due to limited knowledge, we often did not understand why certain herbs were prescribed. I think it would be fantastic if AI could address this. After inputting patient's symptoms and teacher's prescription, AI could immediately provide the theoretical analysis for the prescription[...]. Through step-by-step interaction, I can understand the reasoning process.*" (P2)

Participants also highlighted that young clinicians, particularly those new to clinical work, faced the dual challenges of managing tedious hospital tasks and learning the clinical knowledge they encountered daily. They believed the LLM-based *Shuzhiqihuang* with rapid knowledge retrieval and Q&A capabilities, could significantly enhance their efficiency in resolving clinical inquiries and facilitate more effective learning.

> "*In the outpatient clinic, we often lack time to digest the knowledge. After working at night, finding time to read several books was difficult and overwhelming. If we had this tool, it would be extremely convenient. Even just a few clicks in the clinic to provide a line of thought would be great! We want to learn but lack of such a good platform.*" (P5)

Additionally, some participants (P1, P5, P6, and P8) expressed their desire for AI to be an advanced literature search and analysis tool. They believed that AI could greatly assist clinicians in finding, integrating, and analyzing literature sources, ultimately enhancing their work and study efficiency.

[5]https://www.who.int/publications/i/item/9789240042322

4.4.2 Assisting TCM Training. Assisting TCM inheritance and education was also a significant expectation of our participants, which primarily reflected in two key aspects. On the one hand, our participants highlighted several challenges in education for TCM students, such as "*rigid training models*" (P2) and "*disconnect between theoretical education and practical application*" (P1). As P14 noted: "*Schools are producing 'exam robots'. In clinical settings, those students lack decision-making competence and need further training from clinical teachers.*" P2 also confirmed this "exam robot", using specific examples in TCM assessment to explain:

> "*TCM clinicians undergo a biennial assessment, with a rigid question bank. Students can pass the exam by memorizing theoretical knowledge or questions from this bank. However, this type of examination does not foster the ability to translate theoretical knowledge into clinical skills. For instance, if a question states that the patient has a red tongue with a slippery coating, memorizing that this corresponds to damp-heat syndrome might be sufficient for the exam. Yet, in actual clinical practice, students even do not know how to observe the tongue properly.*" (P2)

Therefore, P2 and other participants strongly desired the LLM-empowered tools to enhance the training of TM clinicians. They, for instance, expected AI to "*use the most up-to-date corpus to help simulate a more realistic and diverse training content and environment*" (P2) and "*simulate an immersive, clinically relevant scenario*" (P6). Additionally, they hoped AI could "*use algorithms to generate a vast, non-repetitive question bank*" (P1) and "*promote human-computer cooperation in Q&A*" (P4). In addition, some participants (P1, P5, and P12) also hoped AI could assist in promoting TCM knowledge to the public, particularly in areas like health preservation and healthcare, to facilitate the broader dissemination of TCM practices.

4.4.3 Assisting TCM's knowledge Iteration and Reproduction. From a deeper perspective, our participants strongly expected that AI tools could play a vital role in the inheritance and knowledge reproduction of TCM. As we discussed in Section 4.2.3, TCM was deeply rooted in ancient literature and theory. However, relying on ancient literature challenged the inheritance and development of TCM. Given this, our participants were optimistic that AI tools like *Shuzhiqihuang* could assist in the digital preservation, application, and knowledge reproduction of TCM, thereby aiding in the broader inheritance and application. P1 was in charge of the organization of ancient literature at one TCM university and had been dedicated to building a digital library for TCM for a long time. He considered AI a vital tool in advancing the digital storage and application of TCM literature. As he explained: "*Converting ancient literature through OCR technology into electronic formats, then creating a database of ancient literature. Users could input symptoms, and the system would find matches in the database, linking them to classic ancient prescriptions, which would be extremely valuable*".

In addition, participants also expressed high expectations for AI tools to facilitate the iteration and reproduction of TCM knowledge, particularly LLMs with solid learning capabilities. They hoped AI could learn from the expertise of renowned senior TCM clinicians, ensuring that their clinical experience and knowledge continued to be passed down. As P4 noted: "*once these senior TCM clinicians pass away, AI that assisted them will also become a formidable TCM practitioner, and their clinical experience and knowledge will continue to be inherited*". P14 also shared a strong desire for AI to learn and carry forward his own knowledge, becoming a digital reproduction of himself, which "*can record my knowledge, summarize my clinical experience, and iteratively generate new knowledge for me*". As he explained: "*The current inheriting method is to train students, but students 'come and go'. If an intelligent tool could continuously learn my knowledge and diagnostic thought process, and I could keep training it to enrich its capabilities, that would be very valuable.*"

5 Discussion

In the preceding sections, we presented TM clinicians' experiences and perceptions of AI-CDS technologies (e.g., *Shuzhiqihuang*) in TM clinical settings. Our findings suggested that while AI-CDS technologies undoubtedly have significant potential in promoting the practical development of TM, most participants expressed existing AI-CDS tools such as *Shuzhiqihuang* have yet to demonstrate sufficient practical utility and effectiveness. Echoing prior research that emphasizes the importance of aligning AI systems with the needs of real-world scenarios [36, 52], we argue that the integration of AI into TM clinical settings must be considered critically in the way technologies are implemented and their intended application scenarios. In what follows, we offer a critical discussion of potential issues that may arise during AI integration, in terms of trustworthiness (Section 5.1), knowledge diversity (Section 5.2), knowledge inheritance (Section 5.3), following the practical recommendations for future studies looking into this research area.

5.1 Building TM-Specific AI Models to Enhance Practical Utility and Clinical Trust

Similar to modern clinical settings [126], TM clinicians' trust and willingness to adopt AI technologies is a critical concern when integrating AI into practical TM scenarios. In our study, participants reported that their use of AI-CDS tools in daily clinical practice remained limited. One of the primary reasons was that current AI-CDS systems had not yet demonstrated the level of professionalism and accuracy required in real-world clinical settings (as discussed in Section 4.1.2). The nontransparent and unexplainable reasoning process further influenced clinicians' trust to AI-generated recommendations. These point to two main research directions in AI for healthcare: improving model accuracy and enhancing interpretability.

In TM settings, various technical efforts have been made to improve model performance and usability, such as prescription pattern discovering algorithms [63, 87, 119], disease prevention [141], medication usage assistance [56], and drug recommendation algorithms [59, 114, 128]. Based on these technologies, various systems and tools have been developed and implemented, such as TCMISS [103], Ancient and Modern Medical Case Cloud Platform [69], Dajing TCM [17], etc. However, according to our participants, these tools had not yet met the expected functions in practice, with one key reason that these technologies primarily follow modern medicine's reasoning logic and scenario requirements [65, 92].

In fact, there were significant differences between TM and modern medicine in terms of knowledge generation methods, reasoning processes, and scenario requirements. Specifically, clinical decision-making in modern medicine was often based on standardized guidelines [35], with unified clinical terminology and structured documentation, as well as standardized diagnostic processes and metrics. In contrast, TM was experiential [44]. Its core theory of "syndrome differentiation and treatment" [92] led to a nonlinear, complex, and multiphase decision-making process, incorporating the process that was difficult to standardize. As a result, technologies adapted from modern medicine often lack customization to TM's unique features, which limits their effectiveness and usability in practical TM settings. We therefore recommend that future AI for TM should be tailored to the unique features of TM. Building on our findings, we propose the following strategies for customizing AI technologies to better align with TM's unique features and needs.

Socio-Contextual Information Integrated Multi-Stage Reasoning. According to our participants as well as existing TM literature, TM reasoning process is a socio-contextual information-integrated multi-stage process (as shown in Fig. 6b). This process is far more intricate than modern medicine, which primarily relies on extracting diagnostic patterns from standardized data. We therefore suggest that AI-CDS technologies for TM could leverage technologies (e.g., graph neural networks [82, 121] and graph contrastive learning methods [88, 130]) to accurately represent the core entities of TM and capture their high-order relationships, enabling support for multi-stage, complex reasoning processes. Meanwhile, cross-modal data fusion techniques [138] and temporal sequence modeling approaches [10, 139] could be employed to integrate diverse socio-contextual, spatio-temporal data, thereby improving the accuracy and robustness of clinical reasoning.

Extracting Knowledge and Patterns from Multi-source, High-quality Datasets. All of our participants emphasized the significance of the established theoretical foundation of TM, which has been developed over thousands of years and forms the core of its current practice. This contrasts with modern medicine's paradigm of deriving new knowledge and patterns from large datasets. Therefore, AI-empowered models for TM should focus on extracting insights and knowledge firstly from classic TM prescriptions, high-quality datasets rooted in existing theoretical knowledge, and then developing new knowledge from vast amounts of data. Specifically, the dataset should incorporate diverse sources of TM knowledge, such as different schools of thought, regional clinical practices, textbook systems, and medical records from senior TM practitioners, helping preserve the internal pluralism and contested nature of TM. Technically, multi-perspective modeling and heterogeneous data fusion methods [143] could be employed to process and integrate these multi-source data effectively.

Human-AI Collaborative Reasoning and Decision-Making Process. Many AI applications in high-stakes domains such as medicine emphasize the importance of human-AI collaborative decision-making process [86, 111]. Our research highlights the critical role of this collaboration paradigm in the context of TM clinical settings, particularly due to the inherently experiential and empirical nature of TM. We suggest that AI-CDS technologies for TM should support human (expert)-engaged interfaces at each stage of reasoning and decision-making, to enhance their practical utility in real-world TM settings. In particular, the systems should allow for negotiable parameter definitions rather than enforcing rigid standardizations.

5.2 Supporting Clinical Data Work to Preserve Data and Cultural Diversity of TM

In Section 4.3, we discussed that TM data and knowledge were often descriptive, fuzzy, dynamic, and uncertain, making them difficult to quantify or standardize. Specifically, unlike modern medicine, which often relies on measurable indicators obtained through technological tools, TM highly replies on symptoms that are typically perceived through clinicians' "look–listen–question–feel", with subjective and descriptive interpretations, such as "slightly yellow tongue coating" or "floating pulse". Further, TM's diagnostic reasoning process (i.e., syndrome differentiation and treatment) is inherently experience-driven, relying heavily on clinicians' personalized, tacit knowledge, accumulated clinical experience, and theoretical insights. These make it particularly challenging to formalize or translate into standardized formats.

However, the modeling and reasoning processes of AI technologies inherently require structured, quantifiable, and rule-based data and knowledge, which stands in tension with TM's inherent data characteristics. Our participants expressed concerns that, while data standardization is necessary for technological implementation, it often leads to an oversimplification of TM's foundational concepts, as well as its inherently dynamic, context-sensitive, experiential, and personalized syndrome differentiation and treatment process. This not only affects the quality and completeness of TM data and knowledge, but also indirectly contributes to severe data cascade issues [80], i.e., "*compounding events causing negative, downstream effects from data issues – triggered by conventional AI/ML practices that undervalue data quality*". Some participants in our study explicitly pointed out the current state of AI-CDS technologies for TM as "*extracting errors from errors*". Moreover, TM is inherently diverse, with distinct regional practices, theoretical lineages, and diagnostic logic across different schools of thought. The standardized representations through AI systems might marginalize these localized and minority practices, thereby undermining the cultural and theoretical diversity of TM.

It is thus crucial to consider how to better preserve the richness and diversity of data throughout the processes of collection, storage, and standardization and avoid data and cultural biases introduced by over-standardization. An effective solution then is to fully support professional TM practitioners in contributing more specialized and diverse data sources. Yet, clinicians in our study expressed that the collecting and recording of high-quality TM data imposed a significant data workload on them [96]. They often lack the time, energy, or motivation to perform these burdensome data tasks at a high standard, leading to lower-quality data and subsequent data cascades [80]. Therefore, we suggest future research focusing on the challenges related to TM data and data work, carefully considering how to better support practitioners' burdensome data work. Building on our findings, we propose the following design considerations for this issue.

Human-AI Collaborative Empirical Data Recording and Processing. Given the empirical nature of TM's diagnostic and treatment processes, we recommend re-constructing EHR data structure in TM setting to better support efficient, high-quality data recording while reducing the data workload. Specifically, to address the issue of incomplete symptom, syndrome, and therapeutic datasets, which require manual input from clinicians, we suggest implementing AI-empowered models for symptom association, syndrome identification, and therapeutic reasoning. These models need to assist in identifying and selecting relevant symptoms, automatically inferring corresponding syndromes and treatment. Additionally, we propose crowdsourced data collection [38, 97] to gradually enrich symptom databases. Given the experiential and fuzzy nature of TM knowledge, the systems need to support fuzzy and experiential inputs–enabling clinicians to enter vague or unstructured information (e.g., "tongue coating is slightly yellow but not thick, possibly damp-heat") via natural language or diagram-based interfaces, and integrate such inputs into the reasoning process.

Multilayer Human-AI Collaborative Data Annotation. Medical data annotation heavily relies on the expertise of professionals. In TM settings, due to TM's experiential nature, collecting high-quality annotated data requires deeper involvement of highly experienced clinicians, which is very challenging in practice due to the limited availability of experienced clinicians [111]. For addressing this, we suggest leveraging multilayer human-AI hybrid collaborative strategies [132], wherein AI models can continuously and actively learn from expert annotations, progressively improving their annotation performance over time, to support data annotation. Some debate-based methods [89], such as LLM-based multiple agents to simulate different experts [100, 104] for annotating, could also be explored in human-AI hybrid strategies. Meanwhile, we can assign annotation tasks in a multilayered manner based on the expertise of clinicians, thereby maximizing the capabilities of both AI and human practitioners at different levels, and optimizing efficiency, accuracy, and overall annotation quality while reducing costs.

5.3 Scaffolding Critical Reasoning for better TM Inheritance and Education

Lastly, as a medicine with a long history, the effective inheritance and education of TM is always a critical issue. Yet, the complexity and experiential nature of TM pose considerable challenges to TM inheritance and education. Our participants thus expressed great expectations for leveraging AI-empowered technologies to assist in TM inheritance and education. However, despite the great expectations, they also raised concerns of, for instance, students' increasingly relying on AI-generated standardized outputs without critically reflecting on, inductively reasoning, or questioning the underlying diagnostic reasoning, which is one of the most essential components of TM inheritance and education. This could lead to students increasingly focusing on the results (e.g., the recommended prescription) rather than deeply considering "why this syndrome differentiation" and "why this treatment method". Over time, this could undermine the critical thinking and flexibility that are central to TM. Given this, we suggest careful attention to how AI systems are framed and positioned to scaffold students' syndrome differentiation skills through iterative, contextualized reasoning.

Guided Training Design. To scaffold young clinicians' and TM students' syndrome differentiation skills through iterative, contextualized reasoning, we suggest adopting existing educational methods (e.g., "Socratic" question-based interaction method [19, 122], problem-based learning [46]) to enhance students' active thinking in diagnostic and therapeutic plans. Meanwhile, step-by-step guided technologies (e.g.,CoT prompting [116]) could also be used to simulate multi-stage clinical diagnostic scenarios and support young clinicians' and TM students' critical thinking in complex reasoning processes.

Human-Centered Visualization. Due to the complexity of TM reasoning processes, many classic prescriptions lack clear explanations and reasoning details. Additionally, experienced TM clinicians often omit critical intermediate reasoning information when formulating prescriptions, making it challenging for younger clinicians and learners to effectively learn classic and professional theoretical knowledge. In this situation, AI can assist by visualizing the reasoning processes, revealing the inference from symptoms to conclusions. Particularly, we suggest designing diagnostic suggestions from different sources (e.g., classical texts, modern textbooks, and AI reasoning) through multi-perspective presentations, encouraging students' critical thinking.

Construction and Reproduction of TM Theoretical Knowledge. As emphasized by our participants, TM is built upon a theoretical foundation developed over thousands of years, which is central of TM. However, this knowledge is primarily preserved in unstructured ancient texts and lacks systematic and structured characterizing and recording, significantly impacting the inheritance and AI-driven development of TM knowledge. Therefore, rather than focusing solely on discovering knowledge and patterns from vast amounts of data and decision support, we suggest that HCI and AI researchers looking into these research areas first concentrate on exploring the existing theoretical foundations of TM and constructing a comprehensive knowledge system, before moving on to knowledge reproduction.

5.4 Limitation and Future Work

This paper presents a preliminary investigation of TM clinicians' adoption, experience, and perceptions to AI-CDS technologies in practical TM clinical settings. Three limitations should be noted. First, we put our primary focus on TCM. Our findings thus may be limited in their generalizability to a broader range of TM types. Nevertheless, TCM is the most typical TM type [75] and forms the foundation for nearly all TM in Asia [15]. We believe our findings provide HCI and CSCW communities with a valuable case for improving human-centered technology design and deployment in TM domain. Second, we primarily adopted qualitative approaches with 16 TCM clinicians to explore our research questions. While qualitative approaches have been widely employed in CSCW and CHI research (e.g., [95, 98]), and are particularly effective in capturing users' in-depth experiences and perceptions, they are also inherently limited by small sample sizes, which might lead to potential sample selection bias [33] and limited sample diversity [118]. In our study, we practically found substantial variability of practical TM in terms of, for instance, clinician department and digitization level, which greatly influence clinicians' interactions, experiences and perceptions of AI-empowered technology. However, our current sample did not cover all subfields of TCM practice. Moreover, we also found that individual clinicians' AI adoption and digital literacy also play critical roles in shaping their acceptance and perceptions of AI-CDS. Yet, given the sample size of the current study, these aspects were not fully examined in this study, which may limit the generalizability of our findings.

In the future, we plan to expand our research scope by recruiting a broader range of clinicians in terms of TCM specialties, credentials, years of clinical experience, and AI literacy, to gain a more comprehensive understanding of AI-CDS adoption in diverse TCM contexts. Meanwhile, we will improve the design of *Shuzhiqihuang* based on feedback from this study and conduct long-term deployments with clinicians. This will allow us to better capture real-world usage experiences and mitigate the potential influence of AI literacy on their perceptions and interactions with AI systems. We also hope to encourage HCI researchers with different TM cultural backgrounds to join this research area to collaboratively explore the development of HCI-empowered TM.

6 Conclusion

Traditional Medicine (TM), as one of the oldest healthcare practices, has increasingly been utilized as either a primary or complementary therapy. In this study, we explored the potential, challenges, and promise of AI-empowered clinician decision support technologies (AI-CDS) from the perspective of TM clinicians in China. Our findings indicate that, despite the development of numerous AI-CDS technologies and tools, their practical implementation in current TM remains constrained. We identified several challenges specific to integrating AI-CDS into TM due to its distinctive characteristics and the complex data-related issues these features create. By comparing the application of AI-CDS in TM versus modern medicine, we discuss strategies to better enhance the integration and effectiveness of AI in empowering TM practices.

Acknowledgments

We thank Shanghai Science and Technology Innovation Action Plan project (No. 21511104500), NSFC grant (No. 62136002 and 62477014), Ministry of Education Research Joint Fund Project (8091B042239), and Fundamental Research Funds for the Central Universities for funding this research.

References

[1] Stephen R Ali, Thomas D Dobbs, Hayley A Hutchings, and Iain S Whitaker. 2023. Using ChatGPT to write patient clinic letters. *The Lancet Digital Health* 5, 4 (2023), e179–e181.

[2] Leung Yeuk-Lan Alice, GUAN Binghe, CHEN Shuang, CHAN Hoyin, KONG Kawai, LI Wenjung, and SHEN Jiangang. 2021. Artificial intelligence meets traditional Chinese medicine: a bridge to opening the magic box of sphygmopalpation for pulse pattern recognition. *Digital Chinese Medicine* 4, 1 (2021), 1–8.

[3] Majid Asadi-Samani, Mohammad-Taghi Moradi, Leila Mahmoodnia, Shahla Alaei, Fatemeh Asadi-Samani, and Tahra Luther. 2017. Traditional uses of medicinal plants to prevent and treat diabetes; an updated review of ethnobotanical studies in Iran. *Journal of nephropathology* 6, 3 (2017), 118.

[4] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine* (2023).

[5] Zhijie Bao, Wei Chen, Shengze Xiao, Kuang Ren, Jiaao Wu, Cheng Zhong, Jiajie Peng, Xuanjing Huang, and Zhongyu Wei. 2023. DISC-MedLLM: Bridging General Large Language Models and Real-World Medical Consultation. *arXiv preprint arXiv:2308.14346* (2023).

[6] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.

[7] Alexandre Blanco-Gonzalez, Alfonso Cabezon, Alejandro Seco-Gonzalez, Daniel Conde-Torres, Paula Antelo-Riveiro, Angel Pineiro, and Rebeca Garcia-Fandino. 2023. The role of AI in drug discovery: challenges, opportunities, and strategies. *Pharmaceuticals* 16, 6 (2023), 891.

[8] Greg Brockman, Atty Eleti, Elie Georges, Joanne Jang, Logan Kilpatrick, Rachel Lim, Luke Miller, and Michelle Pokrass. 2023. Introducing ChatGPT and Whisper APIs. https://openai.com/blog/introducing-chatgpt-and-whisper-apis.

[9] Andrea Burton, Michael Smith, and Torkel Falkenberg. 2015. Building WHO's global strategy for traditional medicine. *European Journal of Integrative Medicine* 7, 1 (2015), 13–15.
[10] Wanlin Cai, Yuxuan Liang, Xianggen Liu, Jianshuai Feng, and Yuankai Wu. 2024. Msgnet: Learning multi-scale inter-series correlations for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 11141–11149.
[11] C-T Che, V George, TP Ijinu, P Pushpangadan, and K Andrae-Marobela. 2024. Traditional medicine. In *Pharmacognosy*. Elsevier, 11–28.
[12] Zhao Chen, Xiaoyu Zhang, RJ Qiu, Yang Sun, Rui Zheng, Haie Pan, et al. 2021. Application of artificial intelligence in tongue diagnosis of traditional Chinese medicine: a review. *TMR Mod Herb Med* 4, 2 (2021), 14–30.
[13] Zhao Chen, Mengzhu Zhao, Liangzhen You, Rui Zheng, Yin Jiang, Xiaoyu Zhang, Ruijin Qiu, Yang Sun, Haie Pan, Tianmai He, et al. 2022. Developing an artificial intelligence method for screening hepatotoxic compounds in traditional Chinese medicine and Western medicine combination. *Chinese Medicine* 17, 1 (2022), 58.
[14] Ming-Huei Cheng, Ching-Liang Hsieh, Chih-Yu Wang, Chin-Chuan Tsai, and Che-Chang Kuo. 2017. Complementary therapy of traditional Chinese medicine for blood sugar control in a patient with type 1 diabetes. *Complementary therapies in medicine* 30 (2017), 10–13.
[15] Felix Cheung. 2011. TCM: made in China. *Nature* 480, 7378 (2011), S82–S83.
[16] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology* 12, 3 (2017), 297–298.
[17] Nanjing Dajing Traditional Chinese Medicine Information Technology Limited Company. [n. d.]. Dajing TCM. http://www.dajingtcm.com/product
[18] Yiming Cui, Ziqing Yang, and Xin Yao. 2023. Efficient and effective text encoding for chinese llama and alpaca. *arXiv preprint arXiv:2304.08177* (2023).
[19] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't just tell me, ask me: Ai systems that intelligently frame explanations as questions improve human logical discernment accuracy over causal ai explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.
[20] Bleiz Macsen Del Sette, Dawn Carnes, and Charalampos Saitis. 2023. Sound of Care: Towards a Co-Operative AI Digital Pain Companion to Support People with Chronic Primary Pain. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 283–288.
[21] Xianghua Ding, Yanqi Jiang, Xiankang Qin, Yunan Chen, Wenqiang Zhang, and Lizhe Qi. 2019. Reading face, reading health: Exploring face reading technologies for everyday health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[22] Yutao Dou, Yuwei Huang, Xiongjun Zhao, Haitao Zou, Jiandong Shang, Ying Lu, Xiaolin Yang, Jian Xiao, and Shaoliang Peng. 2024. ShennongMGS: An LLM-based chinese medication guidance system. *ACM Transactions on Management Information Systems* (2024).
[23] Pham Truong Duy, Nguyen Minh Thanh, Nguyen Anh Vu, and Ly Le. 2017. A machine learning approach for drug discovery from herbal medicine: Metabolite profiles to Therapeutic effects. In *Proceedings of the 8th International Conference on Computational Systems-Biology and Bioinformatics*. 28–33.
[24] Zachary Englhardt, Richard Li, Dilini Nissanka, Zhihan Zhang, Girish Narayanswamy, Joseph Breda, Xin Liu, Shwetak Patel, and Vikram Iyer. 2024. Exploring and characterizing large language models for embedded system development and debugging. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.
[25] Hayley I Evans, Myeonghan Ryu, Theresa Hsieh, Jiawei Zhou, Kefan Xu, Kenneth W Akers, Andrew M Sherrill, and Rosa I Arriaga. 2024. Using Sensor-Captured Patient-Generated Data to Support Clinical Decision-making in PTSD Therapy. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–28.
[26] Jannatul Ferdush, Mahbuba Begum, and Sakib Tanvir Hossain. 2023. ChatGPT and Clinical Decision Support: Scope, Application, and Limitations. *Annals of Biomedical Engineering* (2023), 1–6.
[27] CN Fokunang, V Ndikum, OY Tabi, RB Jiofack, B Ngameni, NM Guedje, EA Tembe-Fokunang, P Tomkins, Salwa Barkwan, Frederick Kechia, et al. 2011. Traditional medicine: past, present and future research and development prospects and integration in the National Health System of Cameroon. *African journal of traditional, complementary and alternative medicines* 8, 3 (2011).
[28] Daniel Waweru Gakuya, Mitchel Otieno Okumu, Stephen Gitahi Kiama, James Mucunu Mbaria, Peter Karuri Gathumbi, Peter Mbaabu Mathiu, and Joseph Mwanzia Nguta. 2020. Traditional medicine in Kenya: Past and current status, challenges, and the way forward. *Scientific African* 8 (2020), e00360.
[29] Boris A Galitsky. 2024. LLM-Based Personalized Recommendations in Health. (2024).
[30] Narayan Ghiotti, David Clulow, Serene Cheon, Kevin Cui, and Hyo Kang. 2023. Prototyping Kodi: Defining Design Requirements to Develop a Virtual Chat-bot for Autistic Children and Their Caregivers. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 126–131.
[31] Chrisna Gouws and Josias H Hamman. 2018. Recent developments in our understanding of the implications of traditional African medicine on drug metabolism. *Expert Opinion on Drug Metabolism & Toxicology* 14, 2 (2018),

161–168.
[32] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O Anatole Von Lilienfeld, Klaus-Robert Muller, and Alexandre Tkatchenko. 2015. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *The journal of physical chemistry letters* 6, 12 (2015), 2326–2331.
[33] James J Heckman. 1979. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* (1979), 153–161. https://doi.org/10.2307/1912352
[34] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 17–24.
[35] Raja Rina Raja Ikram, Mohd Khanapi Abd Ghani, and Noraswaliza Abdullah. 2015. An analysis of application of health informatics in Traditional Medicine: A review of four Traditional Medicine Systems. *International journal of medical informatics* 84, 11 (2015), 988–996.
[36] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C Ahn, Thomas H McCoy, Roy H Perlis, Finale Doshi-Velez, and Krzysztof Z Gajos. 2021. Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–14.
[37] Peter Bai James, Halimatu Kamara, Abdulai Jawo Bah, Amie Steel, and Jon Wardle. 2018. Herbal medicine use among hypertensive patients attending public and private health facilities in Freetown Sierra Leone. *Complementary therapies in clinical practice* 31 (2018), 7–15.
[38] Yu Jiang, Yuling Sun, Jing Yang, Xin Lin, and Liang He. 2018. Enabling uneven task difficulty in micro-task crowdsourcing. In *Proceedings of the 2018 ACM International Conference on Supporting Group Work*. 12–21.
[39] Yuanyuan Jin, Wendi Ji, Yao Shi, Xiaoling Wang, and Xiaochun Yang. 2023. Meta-path guided graph attention network for explainable herb recommendation. *Health Information Science and Systems* 11, 1 (2023), 5.
[40] Yuanyuan Jin, Wei Zhang, Xiangnan He, Xinyu Wang, and Xiaoling Wang. 2020. Syndrome-aware herb recommendation with multi-graph convolution network. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 145–156.
[41] Yanlan Kang, Yang Chang, Jiyuan Fu, Yan Wang, Haofen Wang, and Wenqiang Zhang. 2023. CMLM-ZhongJing: Large Language Model is Good Story Listener. https://github.com/pariskang/CMLM-ZhongJing.
[42] Kumar Kashyp, Arup Kumar Das, Arvind Kumar Bhardwaj, Gourisankar Roymahapatra, Anjana Ghosh, Milan Hiat, Ritesh Jain, et al. 2023. Phytochemical analysis of careya arborea roxb. root extracts: a qualitative analytical approach. *ES General* 1, 4 (2023), 959.
[43] Majeed Kazemitabaar, Runlong Ye, Xiaoning Wang, Austin Zachary Henley, Paul Denny, Michelle Craig, and Tovi Grossman. 2024. Codeaid: Evaluating a classroom deployment of an llm-based programming assistant that balances student and educator needs. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.
[44] Chen Keji and XU Hao. 2003. The integration of traditional Chinese medicine and Western medicine. *European Review* 11, 2 (2003), 225–235.
[45] Sergey V Kovalchuk, Georgy D Kopanitsa, Ilia V Derevitskii, Georgy A Matveev, and Daria A Savitskaya. 2022. Three-stage intelligent support of clinical decision making for higher trust, validity, and explainability. *Journal of Biomedical Informatics* 127 (2022), 104013.
[46] Ari Krakowski, Eric Greenwald, Timothy Hurt, Brandie Nonnecke, and Matthew Cannady. 2022. Authentic integration of ethics and AI through sociotechnical, problem-based learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 12774–12782.
[47] Elena J Ladas, Stacey Marjerrison, Brijesh Arora, Peter B Hesseling, Roberta Ortiz, Federico Antillon, Shalini Jatia, and Glenn M Afungchwi. 2017. Traditional and complementary medicine in pediatric oncology and low-middle income countries: Recommendations from the International Society of Pediatric Oncology (SIOP), T&CM Collaborative. *Journal of the National Cancer Institute Monographs* 2017, 52 (2017), lgx014.
[48] Lixing Lao, Ling Xu, and Shifen Xu. 2012. Traditional chinese medicine. In *Integrative pediatric oncology*. Springer, 125–135.
[49] Min Hun Lee and Chong Jun Chew. 2023. Understanding the effect of counterfactual explanations on trust and reliance on ai for human-ai collaborative clinical decision making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–22.
[50] Min Hun Lee, Daniel P Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2020. Co-design and evaluation of an intelligent decision support system for stroke rehabilitation assessment. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–27.
[51] Florian Lehmann and Daniel Buschek. 2020. Heartbeats in the wild: a field study exploring ECG biometrics in everyday life. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[52] Jiaoyang Li, Lingxiao Zhou, Yi Zhan, Haifeng Xu, Cheng Zhang, Fei Shan, and Lei Liu. 2022. How does the artificial intelligence-based image-assisted technique help physicians in diagnosis of pulmonary adenocarcinoma? A randomized controlled experiment of multicenter physicians in China. *Journal of the American Medical Informatics Association* 29, 12 (10 2022), 2041–2049. doi:10.1093/jamia/ocac179

[53] Qiang Li, Fan Yang, Lisang Liu, Zhezhou Zheng, Xuejuan Lin, and Qinghai Wu. 2014. Classification of diabetes disease using TCM electronic nose signals and ensemble learning. In *2014 9th International Conference on Computer Science & Education*. IEEE, 507–511.

[54] Zhuying Li, Si Cheng, Zhenhuan Chen, Xin Sun, Jiatong Li, and Ding Ding. 2023. Sleepyflora: supporting sleep sharing and augmentation over a distance for social bonding across time zones. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 76–80.

[55] Baoyan Liu, Xuezhong Zhou, Yinhui Wang, Jingqing Hu, Liyun He, Runshun Zhang, Shibo Chen, and Yufeng Guo. 2012. Data processing and analysis in real-world traditional Chinese medicine clinical data: challenges and approaches. *Statistics in medicine* 31, 7 (2012), 653–660.

[56] Can Liu, Kaijie Sun, Qingqing Zhou, Yuchen Duan, Jianhua Shu, Hongxing Kan, Zongyun Gu, and Jili Hu. 2024. CPMI-ChatGLM: Parameter-efficient fine-tuning ChatGLM with Chinese patent medicine instructions. *Scientific Reports* 14, 1 (2024), 6403.

[57] Chang Liu, Changbo Zhao, Guozheng Li, Fufeng Li, and Zhi Wang. 2013. Computerized color analysis for facial diagnosis in traditional Chinese medicine. In *2013 IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 613–614.

[58] Jiannan Liu, Chenyang Li, Jing Xu, and Huanmei Wu. 2018. A patient-oriented clinical decision support system for CRC risk assessment and preventative care. *BMC medical informatics and decision making* 18 (2018), 45–53.

[59] Zhi Liu, Changyong Luo, Dianzheng Fu, Jun Gui, Zeyu Zheng, Liang Qi, and Haojian Guo. 2022. A novel transfer learning model for traditional herbal medicine prescription generation from unstructured resources and knowledge. *Artificial Intelligence in Medicine* 124 (2022), 102232.

[60] Hui Wen Loh, Chui Ping Ooi, Silvia Seoni, Prabal Datta Barua, Filippo Molinari, and U Rajendra Acharya. 2022. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine* 226 (2022), 107161.

[61] Zhuoran Lu and Ming Yin. 2021. Human reliance on machine learning models when performance feedback is limited: Heuristics and risks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[62] Jiajiong Ma, Guihua Wen, Changjun Wang, and Lijun Jiang. 2019. Complexity perception classification method for tongue constitution recognition. *Artificial intelligence in medicine* 96 (2019), 123–133.

[63] Jingyi Mao, Jiong Zhu, Xin Ma, Peijun Ding, and Huimin Zhang. 2022. Data mining'based research on the medication patterns of Professor Sun Shidao's prescriptions for skin diseases. *Shanghai Journal of Traditional Chinese Medicine* 56, 12-16 (2022). doi:10.16305/j.1007-1334.2022.2110099

[64] Jess McIntosh, Charlie McNeill, Mike Fraser, Frederic Kerber, Markus Löchtefeld, and Antonio Krüger. 2016. EMPress: Practical hand gesture classification with wrist-mounted EMG and pressure sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2332–2342.

[65] Razieh Mirzaeian, Farahnaz Sadoughi, Shahram Tahmasebian, and Morteza Mojahedi. 2019. Progresses and challenges in the traditional medicine information system: A systematic review. *Journal of Pharmacy & Pharmacognosy Research* 7, 4 (2019), 246–259.

[66] Nicole Moore, Muhammad Amith, Ana Neumann, Jane Hamilton, Lu Tang, Lara Savas, and Cui Tao. 2024. Translating motivational interviewing for the HPV vaccine into a computable ontology model for automated AI conversational interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–12.

[67] National Medical Products Administration. 2024. China's TCM institutions provide diagnoses, treatment to 1.28 bln people in 2023. (2024). https://english.nmpa.gov.cn/2024-01/23/c_957811.htm

[68] Gary Nestler. 2002. Traditional chinese medicine. *Medical Clinics* 86, 1 (2002), 63–73.

[69] Institute of Information on Traditional Chinese Medicine. [n. d.]. Ancient and Modern Medical Case Cloud Platform. https://www.yiankb.com/

[70] State Administration of Traditional Chinese Medicine. 2022. Notice from the Office of the State Administration of Traditional Chinese Medicine on Issuing the "Statistical Summary Report on the Development of Traditional Chinese Medicine in 2020". http://www.natcm.gov.cn/guicaisi/gongzuodongtai/2022-01-20/24293.html

[71] OpenAI. 2023. GPT-4 Technical Report. (2023). arXiv:2303.08774 [cs.CL]

[72] World Health Organization et al. 2002. *Traditional medicine: growing needs and potential.* Technical Report. World Health Organization.

[73] World Health Organization et al. 2013. *WHO traditional medicine strategy: 2014-2023.* World Health Organization.

[74] Andreas S Panayides, Amir Amini, Nenad D Filipovic, Ashish Sharma, Sotirios A Tsaftaris, Alistair Young, David Foran, Nhan Do, Spyretta Golemati, Tahsin Kurc, et al. 2020. AI in medical imaging informatics: current challenges

and future directions. *IEEE journal of biomedical and health informatics* 24, 7 (2020), 1837–1857.

[75] Hye-Lim Park, Hun-Soo Lee, Byung-Cheul Shin, Jian-Ping Liu, Qinghua Shang, Hitoshi Yamashita, and Byungmook Lim. 2012. Traditional medicine in China, Korea, and Japan: a brief introduction and comparison. *Evidence-Based Complementary and Alternative Medicine* 2012, 1 (2012), 429103.

[76] Bhushan Patwardhan, Dnyaneshwar Warude, Palpu Pushpangadan, and Narendra Bhatt. 2005. Ayurveda and traditional Chinese medicine: a comparative overview. *Evidence-Based Complementary and Alternative Medicine* 2, 4 (2005), 465–473.

[77] Efrén Pérez Santín, Raquel Rodríguez Solana, Mariano González García, María Del Mar García Suárez, Gerardo David Blanco Díaz, María Dolores Cima Cabal, José Manuel Moreno Rojas, and José Ignacio López Sánchez. 2021. Toxicity prediction based on artificial intelligence: A multidisciplinary overview. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 11, 5 (2021), e1516.

[78] JuanZhi Qi, XinYu Wang, and Tao Yang. 2023. Traditional Chinese Medicine Prescription Recommendation Model Based on Large Language Models and Graph Neural Networks. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 4623–4627.

[79] Niroop Channa Rajashekar, Yeo Eun Shin, Yuan Pu, Sunny Chung, Kisung You, Mauro Giuffre, Colleen E Chan, Theo Saarinen, Allen Hsiao, Jasjeet Sekhon, et al. 2024. Human-Algorithmic Interaction Using a Large Language Model-Augmented Artificial Intelligence Clinical Decision Support System. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.

[80] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[81] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[82] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE transactions on neural networks* 20, 1 (2008), 61–80.

[83] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*. 593–607.

[84] Guillermo Schmeda-Hirschmann, Celeste Vega Gomez, Antonieta Rojas de Arias, Alberto Burgos-Edwards, Jorge Alfonso, Miriam Rolon, Francisco Brusquetti, Flavia Netto, Félix A Urra, and César Cárdenas. 2017. The Paraguayan Rhinella toad venom: Implications in the traditional medicine and proliferation of breast cancer cells. *Journal of ethnopharmacology* 199 (2017), 106–118.

[85] Gloria Ser, Ann Robertson, and Aziz Sheikh. 2014. A qualitative exploration of workarounds related to the implementation of national electronic health records in early adopter mental health hospitals. *PloS one* 9, 1 (2014), e77669.

[86] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2023. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence* 5, 1 (2023), 46–57.

[87] Tang Shihuan and Yang Hongjun. 2013. Review of Study on Traditional Chinese Medicine Medication Regularity. *Chinese Journal of Experimental Traditional Medical Formulae* 19, 5 (2013), 359–363.

[88] Jie Shuai, Kun Zhang, Le Wu, Peijie Sun, Richang Hong, Meng Wang, and Yong Li. 2022. A review-aware graph contrastive learning framework for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*. 1283–1293.

[89] Andries Petrus Smit, Nathan Grinsztajn, Paul Duckworth, Thomas D Barrett, and Arnu Pretorius. 2024. Should we be going MAD? A Look at Multi-Agent Debate Strategies for LLMs. In *Forty-first International Conference on Machine Learning*.

[90] Abayomi Sofowora. 1996. Research on medicinal plants and traditional medicine in Africa. *The Journal of Alternative and Complementary Medicine* 2, 3 (1996), 365–372.

[91] Peipei Song, Jianjun Gao, Norihiro Kokudo, and Wei Tang. 2011. Standardization of traditional Chinese medicine and evaluation of evidence from its clinical practice. *Drug Discoveries & Therapeutics* 5, 6 (2011), 261–265.

[92] Yujuan Song, Bin Zhao, Jun Jia, Xuebing Wang, Sibai Xu, Zhenjing Li, and Xu Fang. 2021. A review on different kinds of artificial intelligence solutions in TCM syndrome differentiation application. *Evidence-Based Complementary and Alternative Med.* 2021, 1 (2021), 6654545.

[93] National Data Bureau State Administration of Traditional Chinese Medicine. 2024. Notice from the State Administration of Traditional Chinese Medicine and National Data Bureau on "Issuing Several Opinions on Promoting the Development of Digital Traditional Chinese Medicine". https://www.gov.cn/zhengce/zhengceku/202408/content_6968519.htm

[94] National Health Commission State Administration of Traditional Chinese Medicine. 2020. Notice of the State Administration of Traditional Chinese Medicine and National Health Commission on Issuing the "Classification and Codes of Traditional Chinese Medicine Diseases and Symptoms" and "Clinical Terminology of Traditional Chinese Medicine Diagnosis and Treatment". https://www.gov.cn/zhengce/zhengceku/2020-11/24/content_5563703.htm
[95] Yuling Sun, Jiaju Chen, Bingsheng Yao, Jiali Liu, Dakuo Wang, Xiaojuan Ma, Yuxuan Lu, Ying Xu, and Liang He. 2024. Exploring Parent's Needs for Children-Centered AI to Support Preschoolers' Interactive Storytelling and Reading Activities. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–25.
[96] Yuling Sun, Xiaojuan Ma, Silvia Lindtner, and Liang He. 2023. Data Work of Frontline Care Workers: Practices, Problems, and Opportunities in the Context of Data-Driven Long-Term Care. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–28.
[97] Yuling Sun, Xiaojuan Ma, Kai Ye, and Liang He. 2022. Investigating Crowdworkers' Identify, Perception and Practices in Micro-Task Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–20.
[98] Yuling Sun, Zhennan Yi, Xiaojuan Ma, Junyan Mao, and Xin Tong. 2024. Technology-Mediated Non-pharmacological Interventions for Dementia: Needs for and Challenges in Professional, Personalized and Multi-Stakeholder Collaborative Interventions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.
[99] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent knowledge graph embedding for effective recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems*. 297–305.
[100] John Chong Min Tan and Mehul Motani. 2023. Large language model (llm) as a system of multiple expert agents: An approach to solve the abstraction and reasoning corpus (arc) challenge. *arXiv preprint arXiv:2310.05146* (2023).
[101] Yang Tan, Zhixing Zhang, Mingchen Li, Fei Pan, Hao Duan, Zijie Huang, Hua Deng, Zhuohang Yu, Chen Yang, Guoyang Shen, et al. 2024. MedChatZH: A tuning LLM for traditional Chinese medicine consultations. *Computers in Biology and Medicine* 172 (2024), 108290.
[102] Jinling Tang, Baoyan Liu, and Kanwen Ma. 2008. Traditional chinese medicine. *The Lancet* 372, 9654 (2008), 1938–1940.
[103] Shihuan Tang, Dan Shen, Peng Lu, and Hongjun Yang. 2015. Advances in researches made via traditional Chinese medicine inheritance support system. *China Journal of Traditional Chinese Medicine and Pharmacy* 30, 2 (2015), 3.
[104] Xiangru Tang, Anni Zou, Zhuosheng Zhang, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. 2023. Medagents: Large language models as collaborators for zero-shot medical reasoning. *arXiv preprint arXiv:2311.10537* (2023).
[105] the State Council the Central Committee of the Communist Party of China. 2019. Opinions of the Central Committee of the Communist Party of China and the State Council on Promoting the Inheritance, Innovation and Development of Traditional Chinese Medicine. https://www.gov.cn/gongbao/content/2019/content_5449644.htm
[106] the State Council. 2016. Notice of the State Council on Issuing the Outline of "the Development Strategy Plan for Traditional Chinese Medicine (2016-2030)". https://www.gov.cn/zhengce/zhengceku/2016-02/26/content_5046678.htm
[107] Jagadisha Thirthalli, Liang Zhou, Kishore Kumar, Jie Gao, Henna Vaid, Huiming Liu, Alex Hankey, Guojun Wang, Bangalore N Gangadhar, Jing-Bao Nie, et al. 2016. Traditional, complementary, and alternative medicine approaches to mental health care and psychological wellbeing in India and China. *The Lancet Psychiatry* 3, 7 (2016), 660–672.
[108] C Michele Thompson. 2015. *Vietnamese traditional medicine: a social history*. Vol. 2. NUS Press.
[109] Norbert L Vecchiato. 2019. Traditional medicine. In *The ecology of health and disease in Ethiopia*. Routledge, 157–178.
[110] Dingyu Wang, Jia Ding, Anlan Sun, Shanggui Liu, Dong Jiang, Nan Li, and JiaKuo Yu. 2023. Artificial intelligence suppression as a strategy to mitigate artificial intelligence automation bias. *Journal of the American Medical Informatics Association* (08 2023), ocad118. doi:10.1093/jamia/ocad118
[111] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI doctor" in rural clinics: Challenges in AI-powered clinical decision support system deployment. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–18.
[112] Haochun Wang, Chi Liu, Nuwa Xi, Zewen Qiang, Sendong Zhao, Bing Qin, and Ting Liu. 2023. Huatuo: Tuning llama model with chinese medical knowledge. *arXiv preprint arXiv:2304.06975* (2023).
[113] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 950–958.
[114] Xinyu Wang, Ying Zhang, Xiaoling Wang, and Jin Chen. 2019. A knowledge graph enhanced topic modeling approach for herb recommendation. In *Database Systems for Advanced Applications: 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22–25, 2019, Proceedings, Part I 24*. Springer, 709–724.
[115] Zhe Wang, Keqian Li, Quanying Ren, Keyu Yao, and Yan Zhu. 2023. Traditional Chinese Medicine Formula Classification Using Large Language Models. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 4647–4654.

[116] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.

[117] Wikipedia. 2024. Traditional medicine. https://en.wikipedia.org/wiki/Traditional_medicine#:~:text=Traditional%20medicine%20(also%20known%20as,the%20era%20of%20modern%20medicine.

[118] Henk Wolda. 1981. Similarity indices, sample size and diversity. *Oecologia* 50 (1981), 296–302.

[119] Jiarui Wu, Shihuan Tang, Weixian Guo, Zhang. Xiaomeng, and Zhang Bing. 2014. Comment on applications of data mining used in studies of heritage of experiences of national medical masters. *China Journal of Chinese Materia Medica* 39, 4 (2014), 614–617.

[120] Weilin Wu, Yannan Wang, and Suli Li. 2022. Application Progress of Data Mining Technology in Inheritance of Famous Old Traditional Chinese Medicine Doctors' Experience. *Chinese Archives of Traditional Chinese Medicine* 40, 115-118 (2022). doi:10.13193/j.issn.1673-7717.2022.04.025

[121] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* 32, 1 (2020), 4–24.

[122] Waterfield Xenophon et al. 1990. Conversations of Socrates. *(No Title)* (1990).

[123] Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang'Anthony' Chen. 2020. CheXplain: enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[124] Qihe Xu, Rudolf Bauer, Bruce M Hendry, Tai-Ping Fan, Zhongzhen Zhao, Pierre Duez, Monique SJ Simmonds, Claudia M Witt, Aiping Lu, Nicola Robinson, et al. 2013. The quest for modernisation of traditional Chinese medicine. *BMC Complementary and Alternative Medicine* 13, 1 (2013), 1–11.

[125] Jian-Jun Yan, Rui Guo, Yi-Qin Wang, Guo-Ping Liu, Hai-Xia Yan, Chun-Ming Xia, Xiaojing Shen, et al. 2014. Objective auscultation of TCM based on wavelet packet fractal dimension and support vector machine. *Evidence-Based Complementary and Alternative Medicine* 2014 (2014).

[126] Qian Yang, Yuexing Hao, Kexin Quan, Stephen Yang, Yiran Zhao, Volodymyr Kuleshov, and Fei Wang. 2023. Harnessing biomedical literature to calibrate clinicians' trust in AI decision support systems. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[127] Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. 2024. Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–35.

[128] Liang Yao, Yin Zhang, Baogang Wei, Wenjin Zhang, and Zhe Jin. 2018. A topic modeling approach for traditional Chinese medicine prescriptions. *IEEE Transactions on Knowledge and Data Engineering* 30, 6 (2018), 1007–1021.

[129] Yuanzhe Yao, Zeheng Wang, Liang Li, Kun Lu, Runyu Liu, Zhiyuan Liu, and Jing Yan. 2019. An Ontology-Based Artificial Intelligence Model for Medicine Side-Effect Prediction: Taking Traditional Chinese Medicine as an Example. *Computational and mathematical methods in medicine* 2019, 1 (2019), 8617503.

[130] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph contrastive learning with augmentations. *Advances in neural information processing systems* 33 (2020), 5812–5823.

[131] Wenjun Yu, Mingyue Ma, Xuemei Chen, Jiayu Min, Lingru Li, Yanfei Zheng, Yingshuai Li, Ji Wang, and Qi Wang. 2017. Traditional Chinese medicine and constitutional medicine in China, Japan and Korea: a comparative study. *The American Journal of Chinese Medicine* 45, 01 (2017), 1–12.

[132] Wenjing Yue, Xiaoling Wang, Wei Zhu, Ming Guan, Huanran Zheng, Pengfei Wang, Changzhi Sun, and Xin Ma. 2024. TCMBench: A Comprehensive Benchmark for Evaluating Large Language Models in Traditional Chinese Medicine. *arXiv preprint arXiv:2406.01126* (2024).

[133] Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. 2022. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414* (2022).

[134] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 353–362.

[135] Hongbo Zhang, Junying Chen, Feng Jiang, Fei Yu, Zhihong Chen, Jianquan Li, Guiming Chen, Xiangbo Wu, Zhiyi Zhang, Qingying Xiao, et al. 2023. HuatuoGPT, towards Taming Language Model to Be a Doctor. *arXiv preprint arXiv:2305.15075* (2023).

[136] Heyi Zhang, Xin Wang, Zhaopeng Meng, Yongzhe Jia, and Dawei Xu. 2024. Qibo: A Large Language Model for Traditional Chinese Medicine. *arXiv preprint arXiv:2403.16056* (2024).

[137] Qi Zhang, Jianhang Zhou, and Bob Zhang. 2021. Computational traditional Chinese medicine diagnosis: a literature survey. *Computers in Biology and Medicine* 133 (2021), 104358.

[138] Fei Zhao, Chengcui Zhang, and Baocheng Geng. 2024. Deep multimodal data fusion. *ACM computing surveys* 56, 9 (2024), 1–36.

[139] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 11106–11115.

[140] Xingzhi Zhou, Xin Dong, Chunhao Li, Yuning Bai, Yulong Xu, Ka Chun Cheung, Simon See, Xinpeng Song, Runshun Zhang, Xuezhong Zhou, et al. 2024. TCM-FTP: Fine-Tuning Large Language Models for Herbal Prescription Prediction. *arXiv preprint arXiv:2407.10510* (2024).

[141] Zongzhen Zhou, Tao Yang, and Kongfa Hu. 2023. Traditional chinese medicine epidemic prevention and treatment question-answering model based on llms. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 4755–4760.

[142] George Zhu, F Musumecci, Peter Byrne, Deepti Gupta, and Ekta Gupta. 2017. Role of traditional herbal medicine in the treatment of advanced hepatocellular carcinoma (HCC: past and future ongoing. *Advance Pharmaceutical Journal* 2, 3 (2017), 115–120.

[143] Junhao Zhu, Yuren Mao, Lu Chen, Congcong Ge, Ziheng Wei, and Yunjun Gao. 2024. FusionQuery: On-demand Fusion Queries over Multi-source Heterogeneous Data. *Proceedings of the VLDB Endowment* 17, 6 (2024), 1337–1349.

A Interview Protocol

Stage 1: Understanding Participants' Background and Current Use of AI Technologies

- Could you briefly introduce yourself, including your clinical background, years of clinical experience, medical specialization or area of expertise, primary department or unit you work in, and type of your affiliated institution (e.g., public hospital, private clinic, integrative medicine center, etc.).
- Could you briefly describe your day-to-day clinical work and the main challenges you experienced?
- What kinds of systems or platforms have you used? Using purpose and experience.
- Have you ever used any AI-assisted tools in your practical clinical practice? If yes, what specific tools have you used, in what scenarios, and how do you experience to these tools.
- How would you evaluate existing AI tools in clinical settings? What features or functions are helpful, and what problems or limitations might they encounter?

Stage 2: Tool Demonstration and Evaluation

- Introduce *Shuzhiqihuang* to participants, including its design goal, three main features, and technical principle.
- Invite participants to evaluate the prescription recommendation feature through inputting symptoms and generating recommended results, and rate the result in terms of accuracy, professionalism, usefulness, trustworthiness, and willingness to use. We pre-established six sets of patient symptoms focusing on common health issues, including 1) fever, sweating, aversion to wind, headache, floating and slow pulse; 2) Heaviness in the body, fullness in the abdomen, bowel sounds, vomiting; 3) Restlessness, feeling hot but not cold, dry tongue, bitter mouth, yellow and red urine; 4) Female patient, prolonged menstruation, heavy flow, pale color, thin texture, fatigue, shortness of breath, pale face, pale tongue, thin coating, slow and weak pulse; 5) What is the role of Ge Gen (Pueraria) in Gui Zhi Jia Ge Gen Tang? How does Gui Zhi Jia Ge Gen Tang differ from the basic Gui Zhi Tang? 6) A 75-year-old patient has swelling in the legs and ankles, normal sleep but loose stools, along with hypertension. How should this be treated? Participants could also randomly pose questions based on their expertise.
- Invite participants to evaluate LLM-based health question-answering features through inputting symptoms and generating recommended answers, and rate the result in terms of accuracy, professionalism, usefulness, trustworthiness, and willingness to use.

Stage 3: Experience and Perception to *Shuzhiqihuang* as well as AI-empowered Technologies

- What is your overall impression of this kind of AI-empowered clinical tool?
- What roles do you think such systems could play in TM clinical practice?
- What potential application scenarios can you imagine?
- What do you see as the main benefits and challenges for AI-CDS systems in TM settings?
- If the system were to be further improved, which aspects would you most like to see enhanced?
- Are there other types of AI-empowered technologies you would find useful in your daily clinical work?
- Do you have any additional concerns, suggestions, comments, or expectations that we should consider in future system design?

B Development process and Technical Detail of *Shuzhiqihuang*

- **TCM Knowledge Visualization (Fig. 2)**. We constructed a medicine knowledge graph (KG) centered on herbs and symptoms, integrating TCM and Western medicine concepts under the guidance of TCM clinicians, who defined entity categories, relationships, and data sources. The KG was built from two structured TCM knowledge databases, semi-structured data from herbal websites, and 26,360 prescriptions from renowned TCM experts. It includes 16 types of TCM diagnostic entities (e.g., symptoms, herbs, syndromes, therapeutic effects, flavors, toxicity, and compound molecules) and 16 types of diagnostic relationships. The construction process involved entity recognition, relation extraction, and entity alignment techniques. Ultimately, this KG comprised 123,358 triples and 37,114 TCM entities, serving as the foundation for the Clinical-AI Collaborative Explainable Prescription Recommendation module.

- **Clinical-AI Collaborative Explainable Prescription Recommendation (Fig. 3).** This module's explainable prescription recommendation algorithm was developed under TCM clinician supervision to ensure its recommendations align with TCM clinical reasoning process and pharmacological associations. The types, nodes, and relationships in the metapath-based algorithm were designed with expert input. Specifically, based on the principles of "syndrome differentiation and treatment" and "pharmacological molecular theory" of herbs, we structured the TCM decision-making process into three metapaths and integrated graph neural networks (GNNs) and attention mechanisms for herb recommendation. This approach achieved state-of-the-art (SOTA) performance in prescription recommendation tasks, with detailed quantitative evaluation results provided in Appendix C. Additionally, we designed a Clinical-AI collaborative interface that allows clinicians to adjust recommendation results and incorporated a four-stage clinical diagnostic thought process (Fig. 4) to assist clinicians in understanding the reasoning behind the recommendations.
- **LLM-Based Intelligent Q&A (Fig. 5).** We developed a TCM-specific LLM using prompt tuning, chain-of-thought (CoT) reasoning, and retrieval-augmented generation (RAG). The model was trained on 26 unstructured ancient TCM texts, which were curated, filtered, and verified by TCM clinicians. Tailored for TCM knowledge Q&A, the LLM supports natural conversational interactions, RAG-enhanced health Q&A, and prescription recommendations. It achieved leading accuracy on the Traditional Chinese Medicine Licensing Examination, demonstrating SOTA performance in TCM Q&A tasks. The detailed quantitative evaluation is provided in Appendix D.

C Evaluation of *Shuzhiqihuang*'s Explainable Prescription Recommendation

We use three commonly applied metrics in the prescription recommendation task to evaluate the model's performance: precision, recall, and NDCG. These metrics are defined as follows:

$$Precision@K = \frac{|Top(sc, K) \bigcap hc|}{K}, \tag{1}$$

$$Recall@K = \frac{|Top(sc, K) \bigcap hc|}{|hc|}, \tag{2}$$

$$NDCG@K = \frac{DCG@K}{IDCG@K}. \tag{3}$$

Here, $Top(sc, K)$ represents the $topK$ herbs with the highest predicted scores for the given symptom set sc. $Precision@K$ measures the proportion of the $topK$ recommended herbs appearing in the prescription, reflecting recommendation accuracy. $Recall@K$ represents the proportion of herbs in the actual prescription that are correctly recommended, indicating recommendation completeness. $NDCG@K$ (Normalized Discounted Cumulative Gain) evaluates the ranking quality of recommended herbs, assigning higher importance to correctly recommended herbs appearing earlier in the list. The recommendation list is limited to a maximum of **20** herbs for evaluation. Final results are obtained by averaging the metrics across all prescriptions in the test set.

We used the widely adopted public dataset TCM [40], which contains classic prescriptions from the *Great Dictionary of Traditional Chinese Medicine Formulas*. For comparison, we included both general recommendation algorithms, including CKE [134], RKGE [99], R-GCN [83], and KGAT [113], as well as prescription recommendation algorithms like HC-KGETM [114] and SMGCN [40]. The comparison results are presented in Table 2.

Table 2. The overall performance comparison of explainable prescription recommendation algorithm. Bold text indicates the best performance, while underlining indicates suboptimal performance. p and r are the abbreviations of precision and recall respectively.

Model	p@5	p@10	p@20	r@5	r@10	r@20	ndcg@5	ndcg@10	ndcg@20
HC-KGETM	0.2783	0.2197	0.1626	0.1959	0.3072	0.4523	0.3717	0.4491	0.5501
CKE	0.2692	0.2161	0.1609	0.1901	0.3063	0.4509	0.3636	0.4443	0.5490
RKGE	0.2719	0.2159	0.1584	0.1928	0.3028	0.4401	0.3681	0.4462	0.5469
R-GCN	0.2910	0.2302	0.1682	0.2079	0.3250	0.4667	0.3901	0.4684	0.5691
SMGCN	0.2928	0.2295	0.1683	0.2076	0.3245	0.4689	0.3923	0.4687	0.5716
KGAT	0.2926	0.2299	0.1683	0.2083	0.3225	0.4693	0.3927	0.4691	0.5721
Ours	**0.2941**	**0.2330**	**0.1699**	**0.2105**	**0.3292**	**0.4743**	**0.3935**	**0.4717**	**0.5733**
%Improv. by HC-KGETM	5.66%	6.06%	4.48%	7.46%	7.16%	4.87%	5.86%	5.03%	4.22%
%Improv. by KGAT	0.50%	1.37%	0.97%	1.05%	2.08%	1.08%	0.19%	0.55%	0.21%

The results in Table 2 show that, the algorithm we used outperformed all comparative algorithms across all evaluation metrics. Specifically, compared to the knowledge graph-based prescription recommendation algorithm HC-KGETM, MGAT improved Precision@10 by 6.06% and Recall@10 by 7.16%. Additionally, compared to the general knowledge graph neural network model KGAT, MGAT achieved a 1.37% improvement in Precision@10 and a 2.08% in Recall@10. These results demonstrate that the algorithm used in the Clinical-AI Collaborative Explainable Prescription Recommendation module of *Shuzhiqihuang* achieves state-of-the-art performance in the prescription recommendation.

Fig. 9 presents a case of a prescription recommendation, where the intersection between recommended herbs of our algorithm and those in the actual prescription is highlighted in red. In this case, the accuracy of the recommendation reaches 70%, while the ranking accuracy of the recommended herbs is as high as 98%. This means the important herbs are effectively identified and recommended for the physician's reference. Additionally, according to TCM theory, the herbs that were not matched have similar therapeutic effects to those in the actual prescription. For instance, both *ternata pinellia* and t*atarian aster root* possess the effect of *phlegm dissipation*.

D The performance of *Shuzhiqihuang*'s LLM-based Question-Answering

To quantitatively evaluate the performance of the TCM LLM used in *Shuzhiqihuang*, we adopted the current benchmark, TCMBench [132], for evaluating LLMs in TCM. This benchmark is based on *Traditional Chinese Medicine Licensing Examination* questions, measuring model accuracy in multiple-choice questions. It evaluates three types of questions: (1) Type A1/A2: Tests fundamental for TCM knowledge understanding of LLMs. (2) Type A3: Assesses longitudinal reasoning in clinical decision-making of LLMs. Type B1: Examines the ability of LLMs to establish horizontal connections between TCM concepts. For comparison, we selected both general-purpose LLMs,

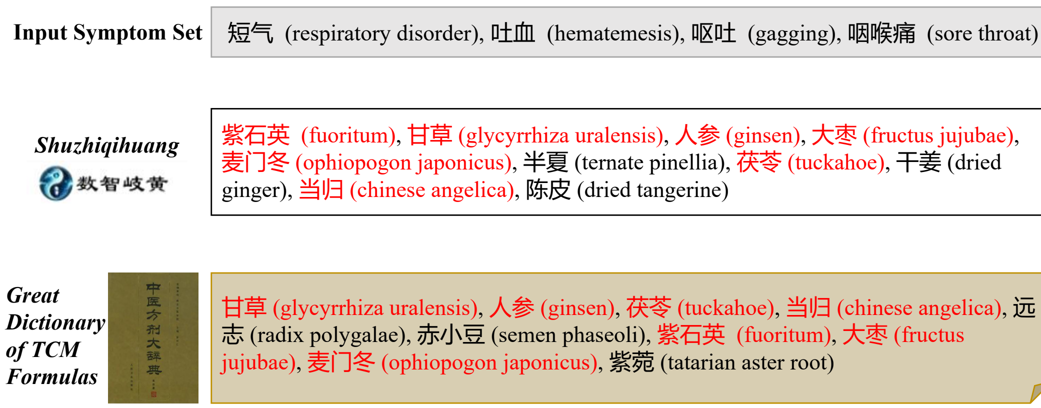


Fig. 9. The case of prescription recommendation.

including GPT-4[71], ChatGPT[8], ChatGLM[133], and Chinese LlaMA[18], as well as domain-specific LLMs, including HuaTuo[112] for Chinese medical applications and ZhongJing-TCM[41] for TCM. The results of the evaluation are presented in Table 3.

Table 3. The accuracy on LLMs in three question types of TCMBench. Bold text indicates the best performance, while underlining indicates suboptimal performance.

LLMs	A1/A2	A3	B1	Total
Chinese LlaMa	0.0969	0.1075	0.1151	0.1089
HuaTuo	0.1944	0.1981	0.1876	0.184
ZhongJing-TCM	0.3537	0.3364	0.2182	0.2695
ChatGLM	0.3613	0.4595	0.4568	0.4477
ChatGPT	0.4510	0.4657	0.4444	0.4398
GPT-4	<u>0.5819</u>	<u>0.6231</u>	<u>0.6011</u>	<u>0.5986</u>
Ours	**0.7444**	**0.7150**	**0.6410**	**0.6799**
% Improve. by GPT-4	27.93 %	14.75%	13.11%	13.58%

The results in Table 3 show that the TCM LLM used in the LLM-Based Intelligent Q&A module of *Shuzhiqihuang* outperformed all comparative algorithms across TCMBench. It achieved an accuracy of 67.99% on TCMBench, outperforming GPT-4 by 13.58% overall. Specifically, accuracy improved by 27.93% for A1/A2 type of questions, 14.75% for A3 type of questions, and 13.11% for B1 type of questions. Compared to the current TCM-specific LLM ZhongJing-TCM, our model achieved nearly 2.5 times higher accuracy. These results demonstrate that our used model significantly reduces the knowledge gap in TCM, enhancing basic comprehension, clinical reasoning, and interdisciplinary knowledge integration capabilities.

Received October 2024; revised April 2025; accepted August 2025