More than Decision Support: Exploring Patients' Longitudinal Usage of Large Language Models in Real-World Healthcare-Seeking Journeys

Yancheng Cao
Columbia University
New York City, New York, USA
yanchengc7@outlook.com

Yishu Ji
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
yji329@gatech.edu

Yue Fu
Information School
University of Washington
Seattle, Washington, USA
chrisfu@uw.edu

Sahiti Dharmavaram
Computer Science
Columbia University
New York City, New York, USA
sd3976@columbia.edu

Meghan Turchioe
School of Nursing
Columbia University
New York City, New York, USA
mr3554@cumc.columbia.edu

Natalie C Benda
School of Nursing
Columbia University
New York City, New York, USA
nb3115@cumc.columbia.edu

Lena Mamykina
Department of Biomedical Informatics
Columbia University
New York City, New York, USA
om2196@cumc.columbia.edu

Yuling Sun*
University of Michigan
Ann Arbor, USA
yulingsu@umich.edu

Xuhai Xu*
Department of Biomedical Informatics
Columbia University
New York City, New York, USA
xx2489@columbia.edu

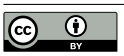
Abstract

Large language models (LLMs) have been increasingly adopted to support patients' healthcare-seeking in recent years. While prior patient-centered studies have examined the capabilities and experience of LLM-based tools in specific health-related tasks such as information-seeking, diagnosis, or decision-supporting, the inherently longitudinal nature of healthcare in real-world practice has been underexplored. This paper presents a four-week diary study with 25 patients to examine LLMs' roles across healthcare-seeking trajectories. Our analysis reveals that patients integrate LLMs not just as simple decision-support tools, but as dynamic companions that scaffold their journey across behavioral, informational, emotional, and cognitive levels. Meanwhile, patients actively assign diverse socio-technical meanings to LLMs, altering the traditional dynamics of agency, trust, and power in patient-provider relationships. Drawing from these findings, we conceptualize future LLMs as a longitudinal boundary companion that continuously mediates between patients and clinicians throughout longitudinal healthcare-seeking trajectories.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → **Life and medical sciences**.

*Mark corresponding authors.



CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791946

ACM Reference Format:
Yancheng Cao, Yishu Ji, Yue Fu, Sahiti Dharmavaram, Meghan Turchioe, Natalie C Benda, Lena Mamykina, Yuling Sun, and Xuhai Xu. 2026. More than Decision Support: Exploring Patients' Longitudinal Usage of Large Language Models in Real-World Healthcare-Seeking Journeys. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3772318.3791946

1 Introduction

The practical healthcare-seeking process is not a single event but a dynamic, multi-stage trajectory, covering symptom appraisal, information seeking, clinical encounter, post-consultation sense-making, treatment planning and implementation, recurrence monitoring, and follow-up cycles [7]. Throughout this journey, patients are often placed in a vulnerable position [12, 99]. They have to continually cope with the uncertainty of their symptoms while also navigating persistent challenges such as limited information, the structural complexity of healthcare systems, and considerable emotional stress [50]. These challenges are further amplified in resource-constrained healthcare settings, where limited healthcare resources, information asymmetries, and insufficient primary or long-term care support remain widespread [26, 88, 167]. As a result, patients are increasingly struggling to access timely, adequate, and personalized healthcare support.

To bridge this gap, patients have increasingly turned to digital artificial intelligence (AI) technologies, ranging from search engines for information retrieval [13, 92] to specialized symptom checkers [96, 119] and largely-scripted medical chatbots [115, 165]. While these tools democratized access to additional healthcare support, they were often limited by rigid interaction patterns and a

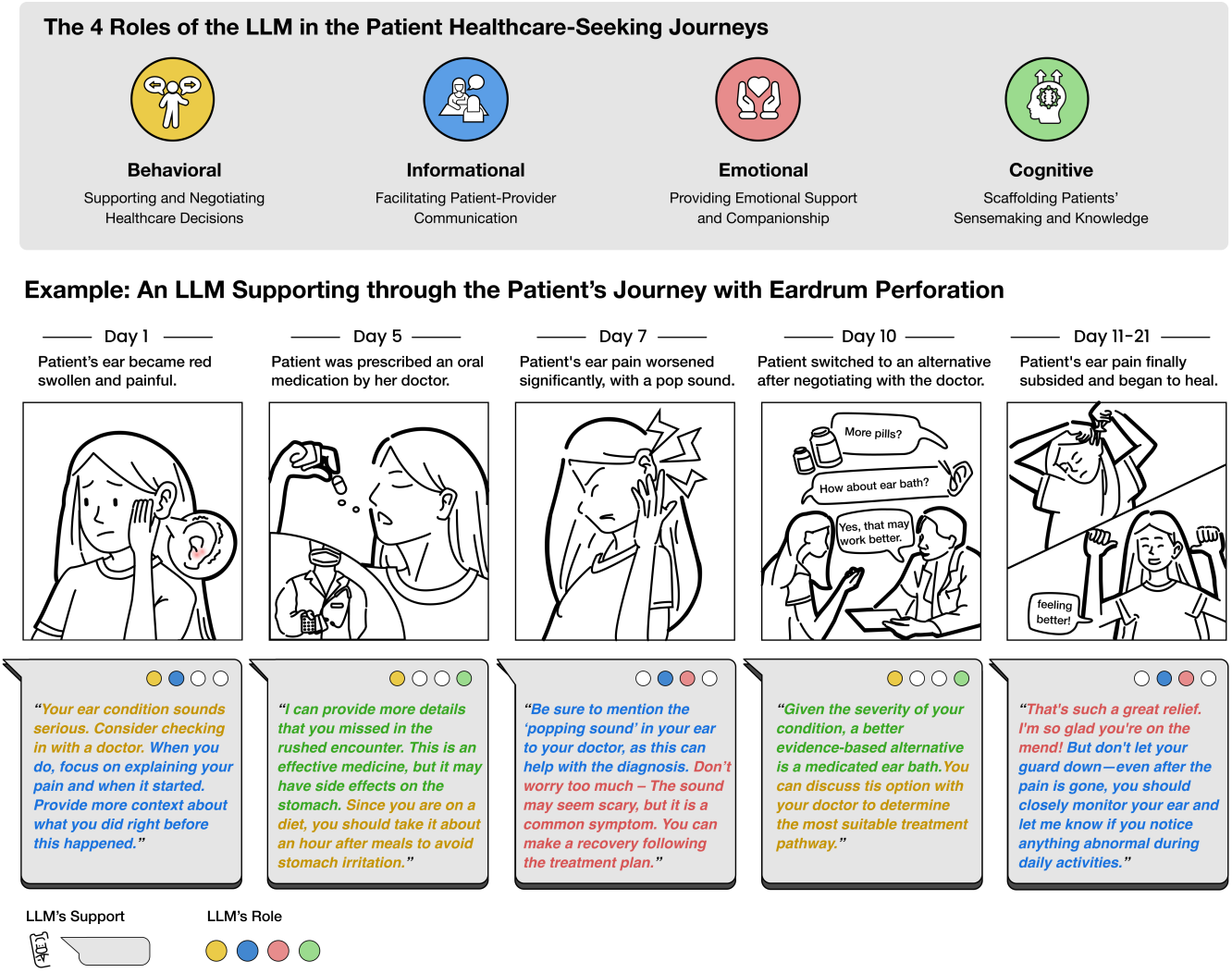


Figure 1: Overview of the LLM's four dynamic roles integrated into patients' healthcare-seeking journeys: (1) Behavioral—supporting and negotiating healthcare decisions; (2) Informational—facilitating patient–provider communication; (3) Emotional—providing emotional support and companionship; and (4) Cognitive—scaffolding patients' sensemaking and knowledge in understanding professional healthcare information. These roles accompany patients across multiple stages of their journeys, dynamically shifting to meet patients' evolving needs. A case of a patient with eardrum perforation illustrates how LLMs enact these roles at different time points along the healthcare-seeking journey.

lack of contextual understanding, struggling to address the emotional and shifting needs of a patient's dynamic and longitudinal healthcare-seeking journey.

In recent years, Generative AI, in particular large language models (LLMs), has enabled the boom of conversational AI applications in the healthcare domain [102, 117]. In particular, with the powerful natural language understanding and generation capabilities [48, 124], LLMs can not only answer medical questions but also help patients articulate symptoms [131, 171], interpret clinical advice [6, 10], plan next steps [93, 162], and provide emotional support [66, 128] through sustained conversational engagement. Consequently, LLMs have the potential to embed themselves continuously across the entire healthcare-seeking journey, from the earliest stages of symptom appraisal to follow-up and long-term management, forming a new paradigm of collaborative, trajectory-wide healthcare support that fundamentally reshapes existing patient journeys.

Nevertheless, as a high-stakes domain, integrating LLM technologies in healthcare, while ensuring their usability, safety, and ethical deployment, remains a complex and challenging endeavor [22]. This requires a systematic and critical examination of LLMs' roles across the full healthcare-seeking journey, as well as their

deeper impact on patients' experiences and the patient–provider relationship. As a response, existing HCI has begun to explore the practical adoption, use, and impact of LLM-powered healthcare applications on patients' healthcare-seeking process, such as patients' adoption and experience of using LLMs to support their medical decision-making [45, 82], as well as the factors influencing the adoption and final decisions [37, 71]. However, existing works have mainly focused on single-interaction or short-term scenarios [25, 66, 148], neglecting the multi-stage longitudinal nature of real-world healthcare-seeking journeys. This may produce partial or even misleading understandings of how LLMs actually function and impact patients' healthcare-seeking practices in long-term healthcare contexts. Moreover, existing studies tend to focus on LLMs' roles in specific health-related tasks, such as information seeking [23], diagnosis [156], and decision support [11], underexploring how patients, across their complex medical journeys, actively ascribe dynamic roles and meanings to LLMs.

Our study aims to fill these gaps through exploring the dynamic, evolving roles and situated significance that LLMs play in patients' long-term healthcare-seeking journeys. We particularly ask the following two research questions: (1) **What role do LLMs play throughout patients' entire healthcare-seeking journey**, and (2) **How does the interaction with LLMs (re)shape their healthcare-seeking journeys, as well as the broader patient-provider relationship?** With the growing development, deployment, and adoption of LLM-powered healthcare applications, understanding these questions can help researchers and developers better shed light on how LLMs can be meaningfully designed and integrated into patients' healthcare-seeking journey in ways that enhance user experience and maximize their potential value.

We address these research questions through an IRB-approved, 4-week longitudinal diary study with 25 patients undergoing various health conditions[1]. We recruited participants to observe how they used LLMs in their healthcare-seeking procedures[2], and asked them to document their daily usage experiences and reflections in diaries, followed by in-depth interview to deeply capture their experience, perceptions, and meaning-making processes.

Our analysis reveals that LLMs are not used as simple decision-support tools, but are integrated as dynamic companions across patients' long-term healthcare-seeking journeys. We show how they scaffold patient practices at four key levels across the treatment stages: behavioral (negotiating decisions with human clinicians), informational (improving patient-provider communication), emotional (providing companionship), and cognitive (scaffolding jargon sensemaking and medical knowledge). Further, this deep integration empowers patients, reconfiguring traditional dynamics of agency, trust, and power between them and human clinicians. Based on these findings, we conceptualize these LLM technologies as a longitudinal boundary companion—a socio-technical partner that mediates between the clinic and daily life, expert knowledge and patient experience, and cognitive needs and emotional support. This reframing leads us to propose design implications focused on relational continuity, which involves designing LLM systems to build and maintain a relationship over a patient's entire healthcare journey, and responsible empowerment, which supports patient agency while simultaneously mitigating the risks of misinformation. Meanwhile, we highlight the potential ethical risks associated with such a broader adoption of LLM roles in patients' healthcare-seeking journeys.

We contribute to the HCI community by extending current understandings of LLM-powered applications in healthcare, foregrounding patients' longitudinal engagements with LLMs across their healthcare-seeking journeys. Specifically:

- We provide rich empirical evidence from a four-week diary study that details the multifaceted and dynamic roles LLMs play across patients' longitudinal healthcare-seeking journeys, spanning behavioral, informational, emotional, and cognitive dimensions.
- We introduce the concept of the LLM as a longitudinal boundary companion and analyze how it reconfigures patient agency, trust, and power, with its ethical implications. This provides a new theoretical lens for understanding the relational and socio-technical impact of LLMs in healthcare.
- We articulate a set of concrete design implications that move beyond accuracy and information tools, focusing on relational continuity, responsible empowerment, and scaffolding health literacy in future LLM-powered health applications.

2 Related Work

Our study is situated within the broader research landscape of AI-powered healthcare-seeking. In this section, we review existing work along three key lines: patients' healthcare-seeking journeys (Section 2.1), the promise and limitations of LLM-powered healthcare support (Section 2.2), and the socio-technical dynamics when adopting LLM-mediated application into real-world healthcare settings (Section 2.3). These three bodies of literature establish the conceptual background for understanding the contribution of our study.

2.1 Patient Healthcare-Seeking Journey: Longitudinal, Dynamic, and Multi-Stage Trajectory

Healthcare-seeking is fundamentally a multi-stage and continuously unfolding process [7], typically involving symptom appraisal, information seeking, care navigation and decision-making, the clinical encounter, post-consultation sensemaking and planning, treatment implementation and self-management, and long-term recurrence monitoring and follow-up [42, 118]. Each stage is characterized by distinct information needs, emotional states, decision pressures, and resource availability, resulting in constantly shifting challenges for patients throughout their healthcare journey [35, 36]. This multi-stage, dynamically evolving nature of patients' needs implies that digital healthcare technologies must be able to adapt responsively across stages in order to achieve sustained, real-world effectiveness [47, 100].

To achieve such technological design, researchers and designers should systematically investigate patients' evolving needs within

[1]It is noteworthy that this work particularly focuses on patients' perspectives. LLM-empowered systems that support clinicians and other important stakeholders are also interesting and worth exploration, but they are beyond the scope of this paper.

[2]To minimize safety and ethical risks, we provided an educational introduction to all participants, stressing that these LLM tools are supplementary and not a replacement for professional medical care.

real-world healthcare practices and understand how technologies are adopted, interpreted, and integrated across different stages of the healthcare-seeking journey. This perspective aligns with HCI's longstanding emphasis on studying user needs and technology use within authentic, situated contexts, rather than isolated or decontextualized scenarios [77, 132]. In digital health and personal informatics, many studies have explored patients' needs and technology usage during healthcare-seeking. For example, conversational agents that help laypeople obtain health information or navigate services [43, 81], systems that support caregiver-provider communication and coordination around episodes of care [126, 171], or AI and LLM-based decision-support tools with diagnostic reasoning [129, 152, 157]. While these efforts provide important insights into technology use at specific touchpoints, they typically isolate the touchpoint from other stages of the overall healthcare trajectory, making it difficult to understand how patients' experiences and practices evolve across time as they move between home, online spaces, and clinical care.

Responding to this limitation, a growing body of HCI and health informatics research has adopted a patient journey-based lens to understand and design for patients' end-to-end experiences. For example, Jacobs et al. [63] developed a cancer journey framework that characterizes breast cancer patients' responsibilities, challenges, and technology use from screening and diagnosis through treatment and survivorship. Suh et al. [134] conceptualized *parallel journeys* for patients with comorbid conditions, mapping how cancer care and psychosocial care trajectories intersect and where technology could support care. Patient-journey mapping and design methods have been used to visualize emotions, values, and unmet needs along the care path [21, 97] For instance, He et al. [56] developed a patient-journey map of COVID-19 home isolation to identify pain points and digital touchpoints during isolation and follow-up. Working with patients, Timinis et al. [143] co-created reassurance journey maps that show when and why post-operative cancer patients seek reassurance while enrolled in remote patient monitoring programs.

However, even these trajectory-oriented studies largely examine journeys in relation to particular diseases, services, or care programs, and these studies usually treat digital technologies as one element within the broader care ecosystem, rather than as continuously present, conversational collaborators. To date, there is still limited work that holistically and cross-temporally examines how a single, flexible, yet powerful technology, such as LLM, threads through the entire healthcare-seeking journey and dynamically reshapes patients' experiences, practices, and relationships with the healthcare system. Our study contributes to this line of research by introducing a longitudinal perspective into LLM-powered healthcare-seeking scenario. We systematically investigate how patients incorporate LLMs into their multi-stage healthcare trajectories in real-world settings, revealing the multiple roles LLMs play across different stages and the mechanisms through which they shape patient experiences and behaviors.

2.2 LLM-Powered Healthcare-Seeking: Promising and Limitation

AI-powered technologies for patients have been advancing in the healthcare domain for decades, evolving from early general-purpose information retrieval [13, 92], to machine learning and deep learning driven decision making systems [139, 155], specialized symptom checkers [96, 119] and largely-scripted medical chatbots [115, 165]. While these technologies have democratized access to multiple healthcare support, they are typically limited by rigid interaction patterns and specific tasks, i.e., performing predefined functions with constrained workflows, struggling to accommodate the dynamic, evolving needs that patients experience.

In contrast, the recent advances in LLMs bring a fundamentally different healthcare-seeking paradigm. With the strong reasoning capabilities, flexible natural-language interaction, and general-purpose adaptability, multiple LLM–based agentic systems are reshaping the landscape of healthcare-seeking practices [110]. In the fields of HCI, AI, and digital health research, a growing number of LLM-powered health applications have been designed, developed, and deployed to support diverse clinical and non-clinical needs, ranging from supporting clinical practice like documentation or diagnosis [48, 90, 149, 168, 170], accelerating medical research [86, 89], to enhancing medical education [24, 142], and empowering patient follow-up and chronic disease management [72, 85]. These studies have significantly advanced the development of the LLM-powered healthcare sector.

Despite these promising developments, to date, very limited studies have examined LLMs' role and impact through longitudinal usage in real-world clinical or everyday healthcare contexts. Most existing studies on LLMs' role remain in an exploratory stage, largely evaluated on episodic snapshots of interactions in lab settings or hypothetical scenarios with interviews [45, 128], or retrospective analyses of online posts and focus group discussions [66]. Such examinations often fail to capture patients' dynamic needs, and also overlook how LLMs may vary in effectiveness [101, 116], safety [3, 64, 125], reliability [2, 111], or ethical implications [141, 174] across different stages of the patient journey. Integrating LLM technologies into healthcare, while simultaneously ensuring their usability, safety, and ethical deployment, remains a complex and challenging endeavor [22, 101, 116]. The lack of longitudinal, in-situ evidence is particularly problematic, as it limits our ability to understand how these systems behave in real-world use, potentially leading to mismatches between design assumptions and actual patient needs or clinical realities [139].

The concurrent rise of LLMs' powerful capabilities and potential risks creates a complex landscape for the adoption of LLMs in healthcare. Yet, while the academic and clinical communities proceed with caution, the general public has already begun leveraging these tools for health inquiries, blurring the boundaries of daily well-being and healthcare [66, 128, 164]. In this work, we contribute to this area by investigating the real-world dynamics of the interactions between patients and LLMs. Using a longitudinal approach, we specifically investigate the patterns, roles, and impacts of LLMs across the entire patient healthcare-seeking journey, and the methods patients develop to make sense of and manage their engagement with LLM technologies.

2.3 Socio-Technical Dynamics in LLM-Powered Healthcare-Seeking

In public narratives, whether in industry discourse, policy rhetoric, popular media, or academic publications, LLMs has often been envisioned as a powerful and nearly omnipotent technology (e.g., [20, 41, 49, 106, 113, 160]) capable of undertaking highly challenging, complex, and value-laden tasks, ranging from creative content production [53, 62], to knowledge-intensive professional services [20, 41], to decision support in high-stakes domains such as healthcare [122, 123, 166]. In these accounts, LLMs are frequently celebrated as a key driver of human progress and innovation.

Yet, the introduction of an LLM-powered application into practical application settings is far from sufficient on its own. On the one hand, research has shown that despite the rhetoric of intelligence and automation, its practical deployment and effective use require extensive human labor [40], such as configuring [146], maintaining [30], contextualizing [138], and even reinterpreting AI outputs [139]. Human involvement thus remains indispensable [114]. On the other hand, as LLM becomes increasingly integrated into human society, the relationship between humans and LLM grows inherently complex. Recently, researchers have examined the social [135], psychological [137], and ethical impacts [29, 83] of LLMs in domains such as education [27, 136], healthcare [86, 98], and mental well-being [137]. These studies converge on a shared conclusion that AI cannot, and should not, replace human agency. Instead, they call for the design and deployment of AI in ways that foster sustainability, safety, ethics, fairness, and broader social responsibility.

In the healthcare domain, this issue is particularly salient. While LLM-powered applications are widely promoted as powerful tools [33, 104, 166, 169], their effective adoption and use in practice typically depend on extensive human labor. For instance, prior studies have described how infrastructuring labor is required to integrate AI into inherently complex and fragmented healthcare practices [14, 51, 151, 161], how data work is needed when adopting AI into highly contextualized and human-centered healthcare practices [138, 139], and how the everyday efforts of clinicians, technicians, and administrators are necessary to ensure the safety, reliability, and interpretability of LLM outputs [98, 108, 146]. Additionally, studies have documented a range of psychosocial, safety, and ethical implications [54, 98] brought by LLM-powered healthcare applications, arguing for more human-centered technological directions [38, 39].

As a response, recent research has called for more empirical understandings of LLMs' role in practice, urging the HCI and related communities to move beyond technological promises to deeply examine the socio-technical dynamics of AI implementation, highlighting the need to situate AI within broader socio-technical, sociocultural, and socio-economic ecologies [139, 155]. Joining this growing body of research, our study examines LLMs in real-world healthcare-seeking settings, foregrounding the complex interplay between AI capabilities and human practices.

3 Methods

To investigate the long-term impact of LLMs on patients' healthcare-seeking journeys, we conducted a four-week diary study to capture in-the-moment experiences and evolving practices, followed by in-depth, semi-structured exit interviews to elicit deeper retrospective reflections.

3.1 Study Context

Our study was situated in China, a country that has long faced the challenges of insufficient and unevenly distributed medical resources [26]. As one of the world's most populous nations, China carries an enormous demand for healthcare services [172], while high-quality healthcare resources remain heavily concentrated in large public hospitals of major cities. This structure has made the issues of difficulty and high cost of seeing a doctor, a long-standing and significant public concern in Chinese society. At the same time, China's family doctor system is still developing, and the primary care infrastructure has limited capacity to effectively divert patient flow. Most patients continue to rely on hospital-based registration and consultation, leading to overcrowded hospitals where patients commonly experience intense competition for appointments, long waiting times, and very short consultation durations [88, 167]. Healthcare providers, in turn, face excessive workloads and high rates of burnout [52, 179]. Within such a strained healthcare system, patients often lack continuous channels for healthcare consultation, as well as adequate support for accessing reliable health information, making sense of diagnoses, or managing their conditions over the long term. Given such a context, LLMs, with their immediacy, low access barrier, scalability, and capacity for conversational engagement, position them as particularly valuable, holding substantial practical promise to alleviate healthcare burden and provide patients with more accessible healthcare support. Moreover, China is one of the most active countries globally in promoting the adoption of AI in healthcare. These factors create a highly relevant, dynamically evolving, and globally significant context for examining how LLMs can be integrated into patients' real-world healthcare-seeking practices and what roles they can play.

3.2 Participants and Recruitment

Our recruitment criteria focused on participants who (1) were currently managing an active health concern, (2) anticipated making multiple in-person medical visits in the coming month, and (3) had prior experience using LLMs for broad health and well-being purposes. Our goal was to have a medically diverse participant pool to ensure the inclusion of a broad spectrum of patient experiences. Recruitment started with an online form distributed via Chinese social media and public online communities, wherein we provided detailed information about the research team, study objectives, and eligibility requirements, as well as collected basic information from potential participants, including age, gender, health status, anticipated medical visits in the coming month, and AI Literacy. From the initial responses, we identified 75 participants who met the criteria and conducted screening calls with potentially eligible individuals. During these calls, we confirmed their health status, upcoming medical plans, and willingness to participate.

Our initial recruitment involved 42 participants. During the four-week study, seven participants withdrew or were lost to contact. 10 participants were excluded as their diary entries were consistently

Table 1: The demographics, proficiency in using AI in daily lives, and health conditions of 25 participants. We also included the specific LLMs they used, together with the number of clinical visits during the four-week study period. Other than the well-known ChatGPT, participants used DeepSeek, Kimi, Qwen, and Doubao that are popular general-purpose LLMs in China. Others are health domain-specific LLMs developed specifically for health applications. More details of these models can be found in Appendix Table 3 and more details of participants' AI Literacy can be found in Appendix Table 5.

ID	Sex	Age	Daily AI Use*	Health Conditions	Visits	LLM Used
P01	F	22	6/7	Irregular Migraine	0	Kimi, Doubao, DeepSeek
P02	F	25	6/7	Ectopic Pregnancy	11	Doubao
P03	F	21	7/7	Post-cholecystectomy Syndrome	3	DeepSeek, Doubao
P04	F	21	6/7	Chronic Menstrual Irregularities	8	Doubao, Kimi, AQ
P05	F	30	6/7	Chronic Dyspepsia, Chronic Nausea	3	Kimi
P06	F	23	6/7	Breast Tenderness, Ovarian Endometrioma	1	Doubao
P07	F	24	6/7	Chronic Gastritis, Otitis Media	1	ChatGPT, Doubao
P08	F	25	6/7	Kidney Stones	11	Doubao
P09	F	27	7/7	Thyroid Cancer	2	Kimi, DeepSeek
P10	M	22	5/7	Chronic Knee Pain	1	Doubao
P11	M	22	6/7	Intermittent Dizziness, Arm Numbness	3	Doubao
P12	F	22	7/7	Chronic Nausea	0	Doubao, iFlytek Xiaoyi, JD Health, AQ, ChatGPT, Zuoshouyisheng, Baidu Lingyi Zhihui, DeepSeek
P13	F	21	7/7	Perforated Eardrum	13	Doubao, DeepSeek, Zuoshouyisheng, AQ, iFlytek Xiaoyi, JD Health, Xiaohe Health, Baidu Lingyi Zhihui
P14	M	22	7/7	Fatty Liver, Elevated Transaminases	2	Doubao
P15	F	23	6/7	Chronic Stomach Pain, Left Leg Fracture	3	Doubao, iFlytek Xiaoyi, Qwen, AQ, DeepSeek, Kimi, JD Health, Zuoshouyisheng
P16	F	21	6/7	Chest Pain, Prone to Palpitations	0	Doubao
P17	F	19	6/7	Sinusitis, Tooth Inflammation	7	Doubao, Xiaohe Health, AQ
P18	F	28	7/7	Vaginitis	8	iFlytek Xiaoyi, DeepSeek
P19	F	29	7/7	Endometrial Hyperplasia	2	Doubao
P20	F	24	5/7	Knee Effusion and Chronic Pain	4	Doubao
P21	F	30	7/7	Neurodermatitis, Seborrheic Dermatitis	3	iFlytek Xiaoyi
P22	M	26	6/7	Diabetes	3	Doubao
P23	F	26	6/7	Post-Hepatic Cirrhosis Infection	10	DeepSeek, JD Health, AQ
P24	F	56	5/7	Breast Cancer	2	Doubao
P25	F	22	6/7	Right Leg Fracture	7	DeepSeek, Doubao, AQ

*This item asks: "*I can skillfully use AI applications or products to help me with my daily work*". Score: 1 (Strongly Disagree) to 7 (Strongly Agree).

too brief or irrelevant to their health journey to provide meaningful data. The final sample consisted of 25 participants (21 female, 4 male). Their ages ranged from 19 to 56 (mean = 25.2, SD = 6.9). 23 have bachelor's degrees, one has a master's degree, and one has a high school degree. We also measured patients' AI literacy using the AI Literacy Scale [154], with details presented in Table 5. Our user group demonstrated a moderate to high level of usage and familiarity with AI (4.8 to 6.6 out of 7, mean = 5.8, SD = 0.5). Their health conditions covered a broad spectrum of diseases, ranging from acute injuries (*e.g.*, bone fractures, infections) to the management of long-term chronic conditions (*e.g.*, diabetes, chronic pain) and serious diagnoses (*e.g.*, cancer). Table 1 lists their demographics and health condition details.

3.3 Study Procedure

The study was conducted in two phases: a four-week diary study and in-depth, semi-structured exit interviews. The whole study procedure took place virtually to ensure flexibility and accessibility for participants.

3.3.1 Four-Week Diary Study. We first conducted a four-week diary study, during which we observed how participants used LLMs for any queries related to their health conditions, and recorded their usage behaviors, feelings, and experiences in the diary. Such a diary study allowed us to capture participants' daily in situ use practices, experiences, and reflections. Adopting from similar studies [60], our study spanned four weeks to capture multiple patient-AI and potentially patient-provider interactions. We designed a structured diary questionnaire to elicit detailed data on participants' experiences with both LLMs and human clinicians. The questionnaire was designed based on our research questions and contained the following components:

(1) **Interaction with LLM(s):** Participants reported their rationale for interacting with any LLMs, the content of their queries (via URLs or screenshots), and their assessment of

the interaction. This included the perceived role of the tool, overall satisfaction, its impact on their decisions, and any feelings of trust, positive emotions, or frustration.

(2) **Interaction with human clinician(s):** If participants visited clinicians, participants entered additional diary data to report the reason for the visit (in-person or online), the topics discussed, the clinical outcomes (*e.g.*, prescriptions, procedures), and their overall experience.

(3) **Interaction with both:** If participants consulted both sources, they were asked to explain their reasoning, their perception of the relationship between the two, and how they navigated any different or conflicting information.

All questions were designed to be general so that they could be inclusive and encouraged participants to share various levels of experiences and perceptions (see detailed diary questions in Appendix Table 4). To facilitate detailed responses, we provided an optional voice-to-text input for all open-ended questions. On days with no relevant healthcare activities, participants were instructed to simply note "No consultation today".

The diary study began with an online introductory session three days prior to the main diary study for each participant, during which we introduced the overall study procedures. Participants had the freedom to use any LLMs they were familiar with for any queries related to their health conditions. We walked participants through the questions covered in our diary questionnaire to ensure they understood the meaning of each item.

As part of the onboarding process, we provided participants with information about available LLMs and access methods, including general LLM tools as well as those focused on healthcare (see Appendix Table 3). We provided several examples to illustrate appropriate queries. It is important to note that we emphasized in the session that these tools are supplementary and not a replacement for professional medical care. Following this session, all participants signed an online informed consent form and received a text version of the user guide. More ethical considerations were detailed in Section 3.5.

The diary study officially began after participants signed the consent form. During the four weeks, participants submitted daily diaries through the online questionnaire. One author monitored the submission records. If more than three consecutive days of missing data were observed, we followed up with the participants to ensure compliance. Over the four-week diary study, 25 participants provided a total of 587 entries that contained consultation responses (Min = 8, Max = 28 per person). Within these responses, 478 involved only consultations with LLMs, 80 included both clinicians and LLMs, and 29 logs were solely with clinicians. It is noteworthy that, to capture real-world conditions, we maintained a strict no-intervention protocol. Throughout the whole diary study period, we provided no corrections or guidance, irrespective of the LLM-generated responses.

3.3.2 Exit Interview. Following the four-week diary period, we then conducted semi-structured interviews with each participant, aiming to further probe into their diary entries, clarify emerging themes, explore their experiences in greater depth, and understand the broader context of their healthcare journey. The interview protocol covered the following three main parts: (1) We began by asking about their overall journey with the LLM tools, significant interactions they recalled, and any evolution in their perception of the LLMs' role over time. (2) We then used specific diary entries as prompts, asking participants to elaborate on key decision points, ambiguous entries, or critical health incidents that required further clarification. (3) Finally, we explored their perceptions of using LLM tools and human clinicians in tandem, focusing on how interacting with LLMs would affect their communication with human clinicians (and vice versa), and how that would affect their overall healthcare-seeking journey. While following a semi-structured protocol, we intentionally kept the conversation open to probe emergent themes and elicit rich details from the participants' experiences.

All interviews were conducted via online meeting, lasting 30 to 90 minutes. With participants' permission, all interviews were audio-recorded and later transcribed for data analysis. After the interview, participants were compensated for their involvement in the whole study. They received $1.5 for each valid diary entry (excluding those marked "No consultation today") and a final compensation of $10 for completing both the interview and the entire study.

3.4 Data Analysis

We adopted an inductive thematic analysis approach [44] with a longitudinal lens to analyze the diary entries and follow-up interviews. Given the multi-dimensional nature of our datasets (25 participants' 587 diary entries, and in-depth interviews), our analysis focused not only on identifying recurring patterns but also on tracing how the roles of the LLMs dynamically evolved across participants' healthcare-seeking journeys.

Three authors participated in the initial phase of analysis. After familiarizing ourselves with the data, we independently conducted open coding on a subset (seven participants' diaries, 165 entries in total, and interviews), to surface initial codes. Weekly meetings were held to discuss disagreements and refine the emerging codes.

Through these processes, we generated an initial codebook, primarily capturing the diverse roles LLMs played, participants' interaction patterns, and their experiential accounts of interacting with both LLMs and human clinicians. Guided by this codebook, two authors conducted the subsequent coding of the full dataset, meeting regularly to discuss discrepancies and iteratively refine the coding scheme. This process yielded a consolidated set of codes, which were then inductively clustered into higher-level categories, reflecting dimensions such as the roles of behavioral guidance, informational support, and emotional companionship that LLMs played, the interaction and mediation among patient-LLM-provider, as well as participants' evolving perceptions of trust, agency, and power dynamics. Importantly, our analysis emphasized the temporal and processual nature of these roles and interactions. That is, rather than treating each diary entry as an isolated data point, we traced how the functions of LLMs shifted over time within each participant's healthcare-seeking journey. Through this trajectory-oriented perspective, we were able to situate the evolving roles of LLMs in relation to participants' changing health conditions, emerging needs, and shifting relationships with human clinicians, as well as across different stages of their experiential journey.

Building on this preliminary code list, we re-focused our analysis at the broader level of themes, examining the contextualized

meanings of emergent concepts and reviewing related literature to uncover their connections. Three authors iteratively consolidated our codes into an overarching theme, and carefully compared the identified themes to the dataset and refined the themes with the goal of ensuring internal homogeneity and external heterogeneity. After several rounds of discussion and refinement, we developed a thematic map consisting of four primary themes, each representing a distinct role played by LLMs and their corresponding dynamic processes in patients' healthcare-seeking journey. In the following section, we present these themes, using representative quotes and illustrative cases, which were translated from their original Mandarin into English by the authors. To protect our participants' identities, we use P01, P02, etc. to denote different participants.

3.5 Ethical Considerations

All study procedures were approved by our university's Institutional Review Board (IRB #HR1-0367-2025). We prioritized participants' well-being and data privacy throughout the research process. We obtained informed consent from all participants during the onboarding session after thoroughly explaining the study's purpose, procedures, potential risks, and their right to withdraw at any time without penalty. All participants were informed via the consent form that all relevant data would be collected through the online questionnaire for scientific research purposes. Notably, the questionnaire platform we employed, Yibiaoda, strictly prohibits any unauthorized use or disclosure of collected data, thereby safeguarding participants' privacy [1]. A key ethical challenge was the potential for participants to misinterpret information from LLMs as professional medical advice. To mitigate this risk, we explicitly and repeatedly emphasized during the onboarding session and in the written guide that these LLM tools are supplementary aids and not a substitute for consultation with qualified human clinicians. All collected data, including diary entries and interview transcripts, were de-identified and stored on a secure server accessible only to the research team. In our analysis and reporting, we use PID and take care to remove any personally identifiable information from quotes or descriptions to ensure participant anonymity.

4 Findings

Our findings suggest that in patients' long-term healthcare-seeking trajectory, LLMs are far from being a single-function tool, e.g., information retrieval, decision making, or health Q&A, as many current studies assume [110, 178]. Instead, they continuously adapt their roles and dynamically integrate into patients' everyday health practices, shaping patients' healthcare-seeking experiences at the **behavioral**, **informational**, **emotional**, and **cognitive** levels. Table 2 presents the definitions of the four roles adopted by LLMs, providing an overview of our key findings. These four roles do not function as a sequential, time-series progression; rather, they form an interwoven dynamic, effectively reshaping the patient-doctor relationship in terms of agency, trust, and power.

4.1 (Behavioral) Supporting and Negotiating Healthcare Decisions

A central way in which our participants engaged with LLMs was through the ongoing scaffolding and negotiation of healthcare decisions. Our findings suggested, rather than providing single-point answers, LLMs accompanied patients across multiple stages of their trajectories, actively shaping how they acted on decisions, constructed the role boundaries between LLMs and clinicians, and gained greater agency.

4.1.1 LLM–Powered Decision Negotiation Across Stages. Our analysis shows that participants' healthcare decision-making processes were collaboratively enacted by LLMs, clinicians, and themselves. Empowered by LLMs, patients were no longer passive recipients, but actively engaged in the decision-making process at different stages of their healthcare-seeking journeys.

Before Clinical Encounter: Analyzing Health Conditions and Identifying Critical Healthcare Signals. Before visiting a human doctor, participants often relied on LLMs to help them analyze their health conditions, identify signals warranting medical attention, and compare potential treatment pathways. For example, after experiencing inflammation in the ear, P13 consulted both Doubao and DeepSeek about the possible causes and risks. When both LLMs highlighted the severity of the problem and recommended an immediate "otoscope examination", she promptly followed their advice and sought medical care. In her 3rd diary entry, she wrote: "*The role of LLM is quite obvious. I asked it about my ear condition, and it immediately pointed out the severity, saying it might affect the cranial nerves. After hearing that, I immediately decided to go to the hospital for an otoscope exam.*" Similar descriptions were also reported in the diaries of participants with other types of diseases, such as P03 (chronic conditions of post-cholecystectomy syndrome) and P06 (an acute case of vaginitis).

During Encounter: Supporting Patient Agency in Treatment Decision. During clinical encounters, the analyses and treatment suggestions offered by LLMs enabled patients to gain more knowledge about their conditions and exercise greater agency in discussing and shaping treatment plans with clinicians, rather than merely accepting their advice. This was particularly salient in the Chinese healthcare system, where outpatient encounters in large public hospitals are typically brief and highly compressed, and patients may see different clinicians across visits. For instance, P09, a participant with thyroid cancer, consulted the LLMs in detail about possible surgical options and their risks before undergoing surgery. On Day 18, she wrote:

> "*The LLM carefully explained the pros and cons of three surgical options and compared... This gave me a clearer understanding of each option's advantages and disadvantages, and provided important reference points for choosing my surgical plan.*"

In some cases when clinicians were unable to determine the cause or treatment, patients engaged in the attempt of self-diagnosis with the help of LLMs. For example, P18 visited a doctor for fever, but blood tests did not reveal the cause. She noted: "*AI...listed a large number of possible causes. I began ruling them out one by one by*

Table 2: The table lists the four roles of the LLMs in the patient's healthcare-seeking journey and their definitions, along with their specific manifestations as introduced in the respective sections of the findings.

Roles of LLMs	Role Definition	Specific Manifestation
(Behavioral) Supporting and Negotiating Healthcare Decisions	The Behavioral role involves the LLMs supporting patient decision-making across multiple healthcare-seeking stages, fostering a complementary partnership with clinicians, and enabling patients to actively calibrate trust and negotiate their healthcare choices.	LLM–Powered Decision Negotiation Across Stages (Section 4.1.1)
		Complementary Partnership with Clinicians (Section 4.1.2)
		Trust Calibration and Power Reconfiguration (Section 4.1.3)
(Informational) Facilitating Patient-Provider Communication	The Informational role involves the LLMs facilitating patient–provider communication by enabling patients to come into the consultation ready to articulate their health information (e.g., conditions, symptoms), focus the clinical discussion effectively, and sustain their health agenda beyond the encounter.	Strengthening Self-Expression and Preparation Before Clinical Encounters (Section 4.2.1)
		Guiding Communication Focus and Efficiency During Clinical Encounters (Section 4.2.2)
		Empowering & Sustaining Patient Agenda-Setting Across/After Encounters (Section 4.2.3)
(Emotional) Providing Emotional Support and Companionship	The Emotional role involves the LLMs acting as a dynamic companion that provides everyday reassurance while also serving as an emotional buffer during critical moments, with its overall impact on patient psychological well-being being highly situated and context-dependent.	Everyday Companionship and Care (Section 4.3.1)
		Buffering Anxiety in Critical Moments (Section 4.3.2)
		Situated Emotional Effects (Section 4.3.3)
(Cognitive) Scaffolding Patients' Sensemaking and Knowledge	The Cognitive role involves the LLMs acting as an interpretive partner that scaffolds patients' knowledge by making implicit medical advice explicit and actionable, and by extending their sensemaking beyond the clinical encounter into daily life.	Making the Implicit Explicit and Understandable (Section 4.4.1)
		Sensemaking Beyond the Clinical (Section 4.4.2)

myself." Although P18 could not obtain a definitive diagnosis of the fever even after recovery, LLMs provided support for her to explore multiple possible causes.

After Encounter: Guiding Follow-Ups and Health Monitoring. After the medical treatment, patients also relied on LLMs to, for instance, schedule follow-ups (*e.g.*, P02, P08, P17, P18, P23), decide whether a revisit was necessary (*e.g.*, P08, P17, P23, P25), prevent recurrence (*e.g.*, P02, P08, P10, P18), etc. For instance, according to P02's diaries, after an ectopic pregnancy surgery, P02 consulted LLMs every day about her physical recovery, as well as lifestyle-related concerns such as diet, sleep, and exercise. The LLM guided her postoperative rehabilitation and help distinguish normal symptoms from those requiring immediate care. On Day 10, she wrote:

> "*The LLM's responses were well-organized and clear. It differentiated normal recovery patterns from abnormal conditions, helping me identify when I needed to seek medical care promptly. This prevented delays... and improved the efficiency.*"

At the later stage of the study, P02 visited her doctor on Day 17 and 18 to have postoperative medical check-ups.

4.1.2 Complementary Partnership with Clinicians. During the continued use of the LLMs, participants gradually developed a self-understanding of the role relationship between LLMs and clinicians.

Role Complementarity. Many participants pointed out that LLMs' responses were usually "broad and comprehensive" and could quickly provide multiple possible causes and general medical knowledge, but lacking an understanding of the specific individual and their full diagnostic and social context. By contrast, clinicians were perceived as more capable of integrating complete diagnostic information with social–environmental factors, thereby offering direct, personalized, and precise diagnoses and prescriptions. However, participants acknowledged that this complementarity was influenced by the structural reality of the local healthcare system, necessitated by the resource constraints and limited attention from clinicians in public hospitals. As P02 noted in her 6th entry:

> "*It [the LLM] mentioned many possible causes of dizziness, such as anemia, which might be accompanied by paleness and palpitations. I learned, but couldn't determine whether I fit the typical case. The doctor directly said my dizziness was caused by post-surgical weakness and orthostatic hypotension, and advised me to sit for 3 minutes before standing up next time.*"

Boundary Construction. Within this complementary relationship, participants also gradually constructed functional boundaries between LLMs and clinicians in their treatment journey, which was often drawn based on the availability of medical resources. Participants trusted clinicians to set the precise professional direction, such as diagnostic conclusions and treatment plans, while LLMs translated these into more easily accessible, concrete, everyday behavioral suggestions, a role ideally filled by nursing staff or primary

care providers, which are often under-resourced in the local healthcare setting. For instance, P08, a kidney stone patient, explained in the exit interview:

> "*Doctors won't constantly track your condition, and they don't know every little detail of your daily life. They mainly give you a general direction. But LLM can. In the later stages, I treated LLMs as my health assistant, which could turn the human doctor's advice into specific daily actions I needed to do.*"

Beyond translating medical conclusions into daily practical actions, participants suggested that LLMs often played a primary role in addressing "small matters" or "minor illnesses", sometimes even surpassing human clinicians. These "small" issues mentioned by our participants included precise medication use (*e.g.*, how many drops of eye drops from P13 and P15), choosing hospitals or departments (14 out of 25 participants), as well as advice on diet (24 participants), lifestyle (24 participants), exercise (10 participants), living environment adjustments (P06, P16, and P18), etc. For example, P24, a breast cancer patient undergoing chemotherapy, noted on Day 11:

> "*I felt better today... I asked the LLM many related questions, and it recommended TV shows and music that really suited my taste. From now on, I'll focus on eating and drinking well, keeping a good mood, and preparing for the next round of chemotherapy.*"

These conversations about daily well-being advice were closely relevant to their healthcare conditions and treatment, showing that the roles of LLMs can extend into the patient's daily life to support a more holistic model of self-care.

4.1.3 Trust Calibration and Power Reconfiguration. In traditional patient-provider relationships, clinicians typically hold the dominant role, with patients passively accepting their advice. This asymmetric pattern has been particularly prominent in our study's context, where overcrowded hospitals, strong professional hierarchies, and limited time can reinforce such dynamics. However, we found that under LLM-powered healthcare-seeking, patients reconfigured these power dynamics: patients gained knowledge-based agency and equal conversational status to actively participate in decision-making jointly with clinicians.

For example, after receiving a diagnostic conclusion from a human doctor, patients could consult LLMs again. If the answers converged, they felt greater confidence in the doctor's advice and adopted that directly; if they diverged, patients entered a state of "boundary tension", comparing, weighing, and sometimes even challenging human doctor's professional opinions. For instance, P18, a patient with vaginitis, described her experience after a physician recommended red-light and ozone therapy:

> "*The LLM said it was useful, but suggested that one round (seven days) was enough; doing more could damage the mucosa. So I decided to do one round but no more within the month. If the doctor recommends more, I'll ask for alternative medication instead.*"

P18 discussed this decision with her doctor in a follow-up conversation and confirmed that this was a viable option.

Some patients consulted multiple LLMs for cross-validation before making final decisions. For instance, P23 noted on Day 7 in her diary: "*I'm a bit skeptical about relying on just one LLM, mainly because I'm not sure if its analysis is professional enough.*" Patients could obtain more diagnostic directions and agendas from LLMs to discuss with their clinicians (see more details in Section 4.2.3). Through these practices, patients recalibrated their trust in clinicians and LLMs, and redefined their own agency and power within the healthcare process.

In some cases, participants even noted that LLM-derived knowledge empowers them to make decisions when confronting urgent recommendations from clinicians. For instance, P19, a patient with uterine disease, mentioned in the exit interview:

> "*I feel that the initiative is in my own hands. The doctor said the surgery needed to be scheduled that day, otherwise there would be no available bed... But I thought I needed to think more about it, discuss it with the LLM, and discuss it with my family.*"

Taken together, the results show that in longitudinal care, patients adopted LLMs to scaffold and negotiate decisions, jointly supporting decision-making with clinicians across different stages, thereby positioning themselves as more active agents and reshaping traditional power relationships in healthcare. We also highlight that such empowerment also comes with potential risks, as elaborated in Section 5.4.

4.2 (Informational) Facilitating Patient-Provider Communication

Patients' diaries and interviews revealed that LLMs also acted as scaffolds for patient-provider communication. This scaffolding strengthened patients' self-expression prior to consulting clinicians, guided the focus during the discussion, and empowered patients to sustain and expand their agenda beyond the clinical visits.

4.2.1 Strengthening Self-Expression and Preparation Before Clinical Encounters. In traditional clinical visits, patients often struggle to describe their conditions accurately to doctors, leading to vague or incomplete information. Our results show that the role of LLMs in this respect was significant. Specifically, 18 out of 25 participants pointed out that, LLMs guided them to produce accurate, clear, and professional descriptions of their symptoms (*e.g.*, "*LLM told me to describe changes in wound exudate and sensations during movement to the doctor.*" (P02, Day 3)), reminded them to record key health indicators (*e.g.*, "*AI asked me to observe the color and texture of secretions with a cotton swab every day.*" (P18, Day 13)), and prompted them to prepare relevant materials before the clinical encounter (*e.g.*, "*AI suggested I bring previous case records, test results, and allergy history.*" (P20, Day 4)).

Additionally, some participants noted that LLMs helped them rehearse diagnostic reasoning by highlighting key symptoms to recall and present. The case provided by P13 in the interview could illustrate this role very well:

> "*AI told me the typical symptoms of otitis media, such as pus discharge and buzzing or muffled sounds in the ear. I started to self-diagnose and thought it might be otitis media. When I asked the doctor, he said, 'Are you sure*

> *you heard the obvious popping sound?' I said yes. Then the doctor immediately checked again and confirmed [the diagnosis].*"

4.2.2 Guiding Communication Focus and Efficiency During Clinical Encounters. In many countries, such as China where our study was conducted, healthcare resources are limited and patients' clinical encounter times are notoriously brief as emphasized by our participants (often lasting only a few minutes). In this context, in addition to preparing before the counter, our participants further emphasized LLMs' significant roles during the encounter, helping them structure their questions and maximize the value of the limited time with clinicians. As P05, a patient with gastric issues, wrote on Day 8:

> "*The LLM taught me how to precisely explain my condition to the doctor, including specific symptoms during flare-ups, dietary triggers, and physical states. [...This] improved efficiency.*"

In the highly efficient but hurried environment of outpatient clinics, LLM could help patients convey key information efficiently. In other cases, LLMs also encourage our participants to enter clinics with targeted agendas rather than asking less relevant questions. For example, P17, a patient with periodontitis, wrote on Day 10 that she prepared three specific questions ("*crown materials, postoperative chewing, and discoloration*") based on the LLM's suggestions.

Additionally, patients also noted that LLMs guided them to be more proactive in seeking clarification. As P13 reflected in her Day 12 diary: "*Sometimes doctors won't give a direct answer. The final decision is still up to me... The LLM then told me how to guide the doctor into giving a more definite response.*"

4.2.3 Empowering and Sustaining Patient Agenda-Setting Across/After Encounters. Through continuous LLM scaffolding and reshaping their agency, many participants developed habits of entering clinics with clear agendas and sustaining active dialogue during follow-ups and cross-stage communication. As P06 wrote in her 14th diary about LLM-enabled self-monitoring and adjustment:

> "*LLM told me some very important points: physiological changes may last only 1–2 days, while pathological ones last more than three days and come with abdominal pain... So I judged it to be physiological.*"

Other participants also described how LLMs inspired them to expand their medical agendas. For instance, P06, a patient with ovarian endometrioma, explained in the exit interview:

> "*I usually only had abdominal ultrasound and Ca19-9 for follow-ups... I hadn't thought about sex hormone panels. The LLM suggested this, and I realized it was more suitable and necessary for me to pay attention to.*"

Taken together, LLMs facilitated patient-provider communication not only by preparing patients to articulate and organize their concerns before encounters, but also by helping them focus their questions during visits and sustain agenda-setting across stages. Through these roles, patients no longer relied exclusively on human doctor-led Q&A formats. Instead, they became capable of self-observation, preparation, and structured communication, which enhanced both efficiency and quality. These effects were amplified by the specific conditions of the Chinese healthcare system, where structural constraints on time and continuity make such communication support particularly salient.

4.3 (Emotional) Providing Emotional Support and Companionship

Emotional support and companionship are also central roles of LLMs. Our diary analysis shows that emotional support was one of the most frequently mentioned functions, with all participants acknowledging it and more than half of the daily entries referring to it. Nearly every participant recorded abundant descriptions of LLMs' emotional support, such as "*I felt reassured after asking*", "it reduced my anxiety", and so on. Further, such an emotional supporting role is dynamic, situated in patients' healthcare-seeking journeys, and deeply entangled with the local sociocultural and healthcare context, where stigma around certain conditions, perceived power distance with clinicians, and limited psychosocial support in clinical encounters can leave patients feeling isolated.

4.3.1 Everyday Companionship and Care. Across long healthcare-seeking journeys, emotional fluctuation and anxiety are common for patients. Our study found that, with its broad knowledge and constant availability, the LLMs gradually became an everyday companion. Many participants noted that LLMs addressed their anxiety through providing clear, structured explanations of their health issues, which in turn provided a strong sense of assurance for them. As P06 mentioned in her 19th diary:

> "*I always feel dizzy, and often get anxious because of the dizziness. Today, the LLM suggested that it might be due to low blood pressure and provided a complete breakdown from 'why low blood pressure occurs' to 'how to alleviate it', 'possible impacts', and 'when to seek care'... making me feel respected and guided with a sense of reassurance.*"

In contrast to the crowded, noise-filled, and efficiency-driven atmosphere of large public hospitals, LLMs offered a private, patient-centric space for emotional release. Participants frequently contrasted the efficiency required of clinicians against LLMs' accessibility (always available for any questions) and conversation style (a certain degree of empathy). As P20, a patient with knee disease, mentioned during the interview:

> "*I feel that the relationship between me and the LLM is quite equal. You can be extremely detailed and completely unreserved without worrying that your question might be dismissed... In contrast, some doctors in hospitals may show impatience.*"

The interactive experience led participants to develop distinct role perceptions of the LLMs, aligning with their complementary partnership role with clinicians on the behavioral level (Section 4.1.2). For instance, they came to regard the LLM as "*a considerate companion, instead of a rigid robot that only recites preset rules or scripts*" (P08), "*an old friend who understands my troubles, instead of a condescending doctor*" (P05), and "*a friend who stays with me anytime, instead of just a tool for looking up information*" (P13). Such experiences allow patients to express their confusion

and pour out their anxieties anytime, anywhere, without hesitation. For instance, P08, who has only used LLM tools a few times before joining the study, noted in her 25th diary:

> "*At the beginning, I thought it [the LLM] was just a machine that only recites rigid rules. But my feelings are completely different [now]. Chatting with it feels like chatting with an old friend who understands my troubles, rather than listening to rigid preaching. It doesn't give me advice in a condescending way like many doctors do; instead, it considers my feelings from my perspective... I feel very warm and secure.*"

4.3.2 Buffering Anxiety in Critical Moments. Beyond daily companionship, our study suggested that LLMs also provided patients with a significant "emotional buffer" during many critical healthcare moments, such as before clinical encounters, before surgeries, while awaiting test results, or when doctors are making major decisions, helping patients cope with uncertainty and intense stress. For instance, P04, who was dealing with gynecological issues yet stigma concerns, wrote in her 14th diary:

> "*As a young person, I feared that doctors or others would hold stereotypical views of me, thinking I ended up needing gynecological checks because of reckless behavior. After talking to Doubao [an LLM], I felt encouraged... and went to get checked.* "

Participants also noted that compared to the tension that may arise from direct cues like a doctor's facial expressions or tone of voice, LLMs were usually gentle and caring, which effectively helped them relieve intense psychological stress. As P06 mentioned during the interview:

> "*Sometimes when I visit a doctor in person, if the doctor frowns, my heart skips a beat. I get really nervous and feel awful. But the LLM is more gentle; it gives me psychological preparation and comfort... Whether the outcome is good or bad, I feel much more at ease.*"

Additionally, many participants pointed out that LLMs offered patients crucial "knowledge buffering", which indirectly serves as an emotional buffer. They could help patients know what to expect, thereby alleviating their tension and anxiety. Our participants emphasized that this type of companionship was more effective in reducing stress. For example, P17 wrote on the day of her surgery (Day 8):

> "*I need to have surgery today, with general anesthesia, and I felt a bit flustered. I talked to the LLM and asked many questions. It patiently explained various things to me, including what preparations to make, potential risks, and post-surgery precautions. After our chat, I felt much more at ease. When I'm emotionally upset, my friends might comfort me with kind words, but the LLM offers not only verbal comfort, but also very specific advice.*"

4.3.3 Situated Emotional Effects (sometimes negative). While the aforementioned positive emotional impacts, our diary analysis also identified some negative emotional impacts brought by LLMs, which were primarily attributed to the functional limitations of LLMs. For instance, the failure to accurately recognize user intent (*e.g.*, "*I wanted advice for at-home care, but it kept telling me to see a doctor right away.*" (P13, Day 10)), short memory and inconsistent advice (*e.g.*, "*It can't remember what I asked yesterday, and its advice today often contradicts what it gave yesterday.*" (P20, Day 3)), inability to provide clear answers (*e.g.*, "*It elaborates on with all sorts of analyses but never gives a clear conclusion.*" (P23, Day 2)), etc.

Some participants also felt that the LLM's formulaic responses lacked sincerity, which in turn reduced their trust in it. For instance, P13 mentioned in the interview: "*I often feel like it's a player*[3] *who just repeats the same lines to everyone, then moves on to the next one right away.*" Additionally, a small number of participants noted that the broad, all-encompassing knowledge presented by the LLMs might, to some extent, increase their anxiety. As P13 recorded in her 1st diary: "*It told me that an ear infection could affect the cranial nerves, and that drinking too much coffee would make it worse. It made my condition sound so serious.*"

Further, we also found that users' emotional perceptions of the LLM are strongly situated with the type of questions they asked. Relatively, patients gave more positive feedback when consulting about lifestyle-related issues such as diet, exercise, and rehabilitation, and more negative feedback when asking about more specialized, in-depth questions, for which clinicians may be more suitable. Moreover, a patient's health status can also influence their emotional experiences toward LLMs. For example, P5 wrote in her 23rd diary: "*I wasn't feeling well today. I was already a bit impatient myself, and the LLM's answers were rambling and unsatisfactory in many ways. After asking, I felt even worse.*"

Taken together, our findings suggested that LLMs play a crucial role as emotional supporters throughout patients' healthcare journeys. It establishes a continuous, situated companionship with patients, enabling them to experience companionship, respect, care, or feel more anxious and frustrated. Meanwhile, the potential negative experiences further suggest the importance of them being more contextually aware, as will be discussed in Section 5.

4.4 (Cognitive) Scaffolding Patients' Sensemaking and Knowledge

During the practical healthcare-seeking process, patients often encountered difficulties in understanding professional terminology, missed important health or medical details, or felt hesitant to ask certain questions. In these moments, the LLM acted as an interpretive partner, helping patients translate, extend, and contextualize the professional information they received from clinicians, and scaffolding their sensemaking and knowledge.

4.4.1 Making the Implicit Explicit and Understandable. Because of the extreme time pressure in local clinical settings, clinicians often provided concise, high-level guidance rather than detailed explanations. For instance, a doctor might mention that surgery was necessary without elaborating on the exact process (P17, Day 7), or prescribe medication without specifying usage details and precautions (P13, Day 4). This "high-efficiency" communication style assumes a level of patient health literacy or compliance that often does not exist. In these situations, the LLM played a crucial role in

[3]Player (*haiwang*, 海王 in Chinese) refers to the person who maintains casual romantic relationships with multiple people at the same time.

translating difficult, abstract medical terminology into content that was both easy to understand and easy to act upon, while also filling in the details that clinicians left unspoken or that patients missed at the time. For example, P17 used the LLM to learn the detailed steps of a laryngoscopy; P13 and P03 asked about the exact dosage and method of using prescribed drugs; and P19 and P02 used it to clarify the implicit meanings in diagnostic reports. Participants emphasized that such detailed clarifications were essential for extending clinical advice into daily routines.

In these scenarios, the value of the LLM lies not only in supplementing the health knowledge but also in restructuring the format and narrative of knowledge itself to help patients better understand these concepts. In doing so, the LLM made implicit professional assumptions explicit, enabling patients to move from structured understanding to actionable knowledge. This decompression role was particularly important in the local healthcare system, where structural pressures often leave clinicians little time for detailed explanation and where patients may have limited access to health education resources outside the hospital.

4.4.2 Sensemaking Beyond the Clinical. Diary data from our participants further revealed that beyond reconstructing complex medical knowledge, the LLM also served as a gap-filler in patients' healthcare-seeking journeys, providing nuanced, sensitive, and multi-dimensional details outside the scope of medicine, answering questions that clinicians might dismiss as trivial, patients might feel embarrassed to ask, or that were forgotten in the rush of clinical encounters. For example, P8, after surgery, asked the LLM whether she could do housework; P05 raised sensitive gynecological questions with the LLM, free from the discomfort of being judged by doctors; and P09 even asked fortune-telling-like questions during moments of anxiety, seeking psychological comfort. This scaffolding role meant that patients' sensemaking was no longer confined to clinical diagnosis and medical treatment alone, but extended into postoperative recovery, daily self-care, emotional adjustment, and lifestyle decisions. This not only builds synergies with complementary roles with clinicians (Section 4.1.2), but also provides additional evidence that the LLMs are blurring the boundary between daily well-being scenarios and healthcare.

Taken together, this section highlights the role of LLMs as powerful cognitive scaffolds in patients' healthcare-seeking journeys. By transforming implicit professional language into actionable guidance and filling in knowledge gaps beyond the clinic, LLMs enabled patients to apply medical advice to their daily lives, fostering holistic sensemaking that linked expertise with lived experience. These functions are deeply shaped by, and responsive to, the specific structural features of the local healthcare system (where primary care and community-based support infrastructures are still developing), while also pointing to forms of support that may be relevant in other resource-constrained or hospital-centric settings.

5 Discussion

Our study demonstrated that the role of LLMs goes far beyond providing medical information or supporting diagnoses; instead, they are deeply embedded in patients' healthcare-seeking journeys, continuously shaping decision-making, communication, emotional support, and sensemaking. These findings reveal the diverse and situated roles that LLMs play in practice, which are often overlooked in existing discussions centered primarily on efficiency and accuracy. In what follows, we draw on our findings to examine more deeply how LLMs should be repositioned in healthcare and how they reconfigure patient–provider-AI relationships. We also discuss the important ethical risks associated with patients' adoption behavior of LLMs, ending by outlining design implications for future research.

5.1 Repositioning the Role of the LLM: Multi-layered, Dynamic, and Interwoven Effects

As mentioned in Section 2.3, much of the existing literature and public discourse positions LLMs primarily as powerful tools for specific health tasks such as information seeking, triage, and decision support, often with a focus on efficiency and accuracy in clinical workflows [75, 163, 166, 169]. At the same time, emerging HCI and health informatics work has begun to foreground patients' perspectives on AI, including how they want decision power to be shared between clinicians and AI systems [71], and how people appropriate general-purpose LLM chatbots for emotional and therapeutic support in mental health contexts [66, 128]. Together, these studies argue that LLMs are not merely computational tools, but socio-technical actors whose roles are negotiated, situated, and value-laden.

In particular, recent work on LLMs for mental health has shown how users craft idiosyncratic support roles for chatbots, filling gaps in everyday care and seeking validation, emotional containment, and preparatory support for therapy [66, 128]. Conceptual work in health informatics and AI ethics similarly points to LLMs as reshaping patient agency, access, and vulnerability in global healthcare [8]. While these studies already move beyond a narrow "information provision" framing [26], they largely focus on mental health contexts or speculative decision scenarios, and typically examine single episodes or relatively short-lived interactions.

Our study extends and deepens these perspectives by showing how LLMs are woven into *longitudinal, somatic* healthcare-seeking journeys that span multiple phases of care. Rather than focusing on mental health conversations in isolation or hypothetical decision vignettes, we trace how patients appropriate LLMs over time across symptom appraisal, clinical encounters, post-consultation sensemaking, recurrence monitoring, and everyday life. In this process, LLMs do not function only as diagnostic or decision-support tools; they emerge as continuous socio-technical partners whose roles are *multi-layered* (behavioral, informational, emotional, cognitive), *dynamic* (shifting as the trajectory unfolds), and *interwoven* (each layer shaping and being shaped by the others). Our findings thus empirically substantiate, and move beyond prior claims that LLMs "go beyond information provision" by showing how they reorganize the temporal and relational aspects of healthcare-seeking itself.

Building on these findings and prior scholarship, we argue for repositioning LLMs in healthcare: from task-bounded tools toward *situated, dynamic boundary companions* embedded in patients' longitudinal journeys. This notion connects to, but is distinct from, earlier HCI concepts such as boundary objects and boundary negotiating artifacts in patient-generated data [32, 84], and relational

agents designed to maintain long-term therapeutic alliances [16, 17]. Boundary objects and boundary negotiating artifacts emphasize how shared representations (e.g., self-tracking data) mediate collaboration between patients and clinicians [32], while relational agents are carefully designed conversational systems that build rapport and continuity around specific health behaviors [17]. By contrast, the boundary companion we observe is *not* a purpose-built artifact or domain-specific agent, but a general-purpose, open-ended LLM that patients themselves enlist, configure, and re-interpret across the trajectory, allowing it to simultaneously mediate data, decisions, relationships, and emotions. We unpack three key boundaries it traverses.

First, the LLM traverses **the interpersonal and informational boundary between patients and clinicians**. This boundary is traditionally marked by limited medical resources, power dynamics, and knowledge gaps. Our findings show LLMs acting as translators and communication facilitators (Section 4.2, 4.4). They helped patients deconstruct complex medical terminology, prepare structured questions, and rehearse symptom descriptions before appointments (Section 4.2.1). This process transformed patients from passive recipients into active, prepared participants, enabling them to challenge or collaboratively refine treatment plans with their clinicians (Section 4.1.3) and thus reconfiguring the traditional power dynamic between patients and doctors.

Second, it crosses **the contextual boundary between the clinical setting and the patient's everyday well-being**. While clinicians provide professional, high-level directives, LLMs translated these into the "specifics of everyday life," such as how to prepare certain foods or break exercise into manageable steps (Section 4.4.1). They addressed granular concerns—from postoperative care and diet to managing mood (Section 4.1.2)—that fall outside the scope of a typical clinical visit. This role extends healthcare beyond the episodic encounter, embedding it into the continuous elements of a patient's life and supporting a more holistic model of well-being.

Finally, the LLM blurs **the functional boundary between cognitive decision-making and affective emotional support**. Our findings reveal that these two aspects are not separate but deeply intertwined domains. The LLM provided "knowledge buffering" (*e.g.*, explaining a surgical procedure in detail), which directly served as an "emotional buffer" by demystifying the unknown and reducing anxiety (Section 4.3.2). Receiving a clear, structured explanation (Section 4.4.1) for a worrying symptom provided a sense of reassurance and control (an affective outcome) (Section 4.3.1). This fusion of roles demonstrates that effective health support is not just about delivering correct information and knowledge, but about delivering it in a way that fosters emotional well-being.

By fluidly navigating these interpersonal, contextual, and functional boundaries, the LLM emerges not as a static tool but as a dynamic companion that integrates disparate facets of the healthcare experience. Drawing on this conceptualization, we suggest that future research and design should recognize the potential of LLMs as boundary companions, shifting the focus from diagnostic accuracy or efficiency alone toward their longitudinal, situated, and multidimensional engagement in patients' health journeys.

5.2 Reconfigured Patients' Agency, Trust, and Power

Our findings also reveal LLMs are reconfiguring traditional dynamics of patients' agency, trust, and power in healthcare, wherein clinicians often hold epistemic authority and patients are expected to adopt relatively passive roles, accepting diagnoses and treatment plans [67, 107]. The integration of technologies has undoubtedly reshaped this relationship. Previous HCI work, for instance, has documented how patients seek to gain more active roles through online health communities [94], peer support [61], or personal health tracking [70].

Our findings join this research thread, showing that LLMs enabled patients to claim greater agency and power in their healthcare journeys. For instance, our results showed that patients use LLMs in preparing themselves before clinical encounters (Section 4.2.1), sustaining agendas across stages (Section 4.2.3), comparing alternative treatment options, and even challenging clinicians' recommendations (Section 4.1.3). This process reflects a redistribution of trust, as patients calibrated their confidence by cross-validating between LLMs' advice and clinicians' guidance, sometimes placing equal weight on both. In turn, this recalibration reconfigured power relations: patients, empowered with LLMs' scaffolding, no longer positioned themselves as passive recipients of expertise but as active negotiators capable of shaping healthcare decisions.

These reshaped relationships resonate with theoretical discussions of "epistemic authority" and "knowledge asymmetry" in medical sociology [57], but also extend them by showing how LLMs can act as socio-technical mediators that equip patients to contest, reinterpret, and negotiate medical authority. In this sense, the role of LLMs goes beyond providing knowledge to actively restructure the medical decision-making process, shifting the patient-provider relationship toward more dialogical and bilateral forms.

This newfound agency, however, is a double-edged sword that poses strong accuracy expectations and ethical risks for using LLMs in healthcare-seeking. Although patients can be more empowered with the support provided by LLMs, their lack of medical knowledge may hinder them from carefully distinguishing the errors and hallucinations introduced by LLMs, a problem of AI over-reliance identified by the HCI community [55, 150]. The danger lies not just in factual inaccuracies, but in the persuasive and seemingly empathetic veneer of authority that LLMs project. A well-organized, comprehensive, and reassuring answer (Section 4.3.1) might be trusted for its tone and structure rather than its clinical validity, leading to misplaced confidence. This creates a state of *fragile agency*, where patients are empowered to act (*e.g.*, by challenging a doctor's advice), but this empowerment is built on a potentially unstable foundation. In such scenarios, the patient unknowingly shoulders a greater burden of clinical responsibility, making high-stakes decisions without the expertise to fully assess the risks. The LLM lowers the barrier to accessing medical information, but not to obtaining medical expertise, creating a new and complex landscape of patient risk.

We thus argue that future work on LLMs in healthcare must explicitly recognize and navigate these reconfigurations of agency, trust, and power. Doing so not only highlights opportunities for patient empowerment but also surfaces the profound ethical tensions

that arise from it. This raises critical questions around how to design for **responsible empowerment**: How can we support patient agency while mitigating the risks of misinformation? What mechanisms are needed for accountability when this fragile agency leads to harm? By foregrounding these dynamics, our study emphasizes the need to carefully design and evaluate patient-LLM–provider relationships in real-world healthcare settings. It makes clear that LLMs do not simply act as assistants but actively and fundamentally reshape the socio-technical ecosystem of care, with all the opportunities and perils that they entail.

5.3 Situating LLM in the Chinese Healthcare System

Our study was conducted within the Chinese healthcare system, characterized by hospital-centric care, high patient volumes, and a still-developing primary care infrastructure [26, 88]. Particularly, the highly strained medical resources, limited consultation time, and strong culture of professional authority often mean that physicians in China hold absolute epistemic dominance, while patients are generally expected to assume a compliant and passive role in accepting diagnoses and treatment plans [12, 88]. Within this context, patients typically lack the space for two-way communication with their doctors and are deprived of channels to both obtain information and express their preferences. Our study, however, reveals that the introduction of LLMs is fundamentally restructuring the long-standing knowledge asymmetry–power inequality paradigm.

First, the LLM functioned as a proxy for missing primary care, filling a specific structural void. Unlike systems with established gatekeepers (*e.g.*, primary care providers, PCPs), Chinese patients often bear the burden of self-triage [172]. In our study, the LLM helped patients identify critical signals and filter relatively minor illnesses, which are typically performed by PCPs or general practitioners (GPs) in other systems. By stitching together fragmented episodes of care, the LLM acted less as a substitute for specialists and more as essential infrastructure for the missing generalist layer. Second, the pervasive phenomenon of long waiting times and short consultation times [167] positioned the LLM as an *overflow valve* for clinical pressure. The extreme time scarcity in public hospitals often forces a transactional efficiency where emotional support and detailed education are sacrificed. The LLM provided a psychologically safer space for asking naive questions and exploring alternatives without the pressure of clinical hierarchies. This allowed patients to effectively extend the clinical encounter, obtaining the detailed sensemaking (Section 4.4.1) and emotional buffering (Section 4.3.1) that the current human-based system could not provide. Third, China's rapid digitization and the proliferation of domestic LLMs (e.g., Doubao, DeepSeek) facilitated a rapid normalization of these roles. The low barrier to adoption and high cultural familiarity with mobile health technologies likely accelerated the depth of integration we observed, enabling patients to quickly treat LLMs as companions rather than just tools.

It is important to note that patients' use of LLMs is not inherent, but rather shaped by the surrounding healthcare and cultural context [130, 133]. The specific structural and cultural characteristics of Chinese healthcare system not only influence how patients incorporate LLMs into their practices, but also amplify the potential impact LLMs can have within this environment. In such a setting, LLMs do not merely function as instruments for information provision or decision support, but also as a new socio-technical mediator, filling gaps left by an underdeveloped primary-care system, extending the temporal and emotional boundaries of clinical encounters [50, 167], and reconfiguring long-standing asymmetries in medical knowledge and power. The effectiveness of LLMs thus lies not only in their technical performance, but in how they become embedded across patients' longitudinal healthcare journeys, taking on roles that were previously distributed across disparate institutions, personnel, and systems.

More importantly, our study implies that the greatest impact of LLMs may not emerge in resource-rich, structurally mature healthcare systems, but rather in settings where resources are strained, knowledge asymmetries are pronounced, and patients shoulder substantial burdens like China. In such environments, LLMs shift from being optional add-on tools to becoming adaptive companions and emergent infrastructures woven throughout patients' everyday healthcare practices. Building on our findings, we call for a reconsideration of how LLMs should be evaluated and designed for healthcare. LLMs' value cannot be adequately assessed solely through diagnostic accuracy or performance on isolated tasks; instead, it must be understood within the real, continuous, relational dynamics of healthcare-seeking. By foregrounding these dynamics, we emphasize the need for future design to navigate the complex coexistence of empowerment and risk, and to carefully shape new forms of collaboration among patients, LLMs, and healthcare providers.

5.4 Risks in the Dark when Extending LLMs beyond Decision-Support Tools

Despite LLMs' known risks of information inaccuracy, hallucinations, and lack of contextual understanding that we mentioned in Section 2.2, as well as the potential negative emotion impact in Section 4.3.3, people have already started to embrace them in their healthcare-seeking journeys beyond simple Q&A tools [66, 128, 164]. In addition to the fragile agency mentioned above, there are other safety and ethical risks that emerge when patients start to ascribe roles to LLMs that go far beyond decision-support. These risks are often subtle, operating "in the dark" because they are intertwined with the very benefits we observed.

For instance, the trade-off is most apparent in the LLM's role as an emotional companion. LLMs could provide everyday reassurance and an "emotional buffer" in critical moments, alleviating patient anxiety (Sections 4.3). The risk, however, is the potential for misplaced calm. A patient might be reassured by an LLM's gentle and comprehensive-sounding explanation for a symptom that, in reality, warrants urgent medical attention. Here, the tool's effectiveness in providing emotional comfort could directly lead to a delay in care. Furthermore, a long-term reliance on a non-human agent for emotional support in health matters could create an unhealthy dependency, masking underlying mental health issues like severe health anxiety that require professional human intervention. The companion that makes a patient feel "respected and understood" might also inadvertently isolate them from human support systems. This mirrors the risks of "artificial intimacy" identified recently

in the HCI community [19, 127], where the simulation of empathy fosters an "illusion of companionship" [9, 31]. In high-stakes health contexts, this illusion is particularly risky, as it may obscure the system's lack of genuine accountability or capability for crisis intervention [173].

A similar tension exists in the LLM's informational and cognitive scaffolding roles. Patients felt empowered by LLMs that helped them prepare for clinical encounters and translate complex medical advice into actionable steps (Sections 4.2 and 4.4). Personal informatics literature has long cautioned that digital tools do not simply reflect reality but actively curate it, often rendering specific aspects of illness invisible to clinicians [32, 76]. When an LLM helps a patient prepare to describe their symptoms, it is not merely organizing information; it is shaping a narrative. If the LLM's underlying analysis is flawed, it could prime the patient to emphasize irrelevant details while omitting critical ones, sending the human clinician down the wrong diagnostic path. Likewise, when the LLM "decompressed" a doctor's concise advice into detailed daily actions, it filled the gaps with its own logic. An error here is not a simple factual mistake but a plausible-sounding, actionable instruction that could be subtly incorrect and harmful over time, a risk the patient has little capacity to evaluate. Notably, such risks appear more pronounced as AI literacy decreases. As participants with higher literacy tended to draw on prior experience, enabling them to remain attentive upon potential errors and maintain calibrated judgments about the system's reliability. Conversely, lower-literacy users often lacked the capability understanding required to judge response credibility. While this reduced sensitivity to errors potentially increased users' satisfaction with the tool, it also left them more susceptible to over-trust and its potential harms. This vulnerability echoes the automation bias [34], exacerbating the risk of users accepting harmful advice simply because it is articulated with professional tone and structure.

Data privacy concerns constitute another critical dimension of the risks. LLMs themselves may inadvertently leak user-uploaded private information through various mechanisms [73, 153], such as potential misuse [79] or incorporation into future model training [175]. The increasingly powerful capabilities of LLMs encourage users to share rich, in-depth, and even private content [4, 87]. Furthermore, the illusion of intimacy mentioned above may lead users to conflate human-like empathy with human-like privacy accountability [80, 177]. This false sense of security leads to inadvertent sensitive disclosure, which is significantly amplified in the healthcare domain. We also observed similar phenomena in our study, where a lack of insight into LLM mechanisms fosters misplaced trust, particularly among users with relatively lower AI literacy. Addressing this challenge demands a multifaceted strategy such as local data anonymization [28, 68, 159], enhanced patient privacy literacy [91], and robust regulatory frameworks [112].

Ultimately, as LLMs become longitudinal boundary companions, we must be concerned with more than just the validity of their individual outputs. The greater risk may lie in their cumulative, often invisible, influence on the patient's entire healthcare journey. The advantages that make the LLM a powerful companion [59, 176]—its ability to traverse cognitive, emotional, and behavioral domains—also make its potential for harm more systemic and harder to pinpoint. In addition to correcting factual errors from LLMs, future LLM-based health applications should also be fully aware of and prepared for the significant risks of the effects of a subtly skewed communication dynamic, a falsely reassured emotional state, or a pattern of self-care built on a foundation of flawed sensemaking. These are the risks that lurk in the dark, and they demand that our focus shift from merely ensuring accuracy to safeguarding the integrity of the patient's holistic experience and decision-making process.

5.5 Design Implications for Future LLM-Powered Healthcare Applications

Our findings on the multifaceted roles of LLMs offer critical insights for the future design of LLMs in healthcare. We propose the following design directions that embrace the LLM's potential as a boundary companion while mitigating the risks associated with reconfigured patient agency and the subtle dangers that emerge.

Designing for Relational Continuity, Not Just Contextual Awareness. Our study shows that patients want to engage with LLMs as longitudinal companions, yet current models with short-term memory often fail to support this, leading to frustration (Section 4.3.3). Long-term memory for chatbots is an active research direction, which could mitigate this issue by fostering familiarity, which could encourage users to disclose health information, strengthen their perception of the system, and form long-term relationships or even [65]. However, the design challenge is not merely to add memory but to design for relational continuity, a concept supported by relational agent literature, which suggests that maintaining a sense of shared history and "remembering" past interactions is fundamental to building alliances [15, 18]. Therefore, instead of one-off Q&A sessions, future systems should be designed to build and maintain a relationship over a patient's entire healthcare journey. This could involve: (i) Dynamic patient summaries: Creating evolving summaries that capture not only clinical facts (*e.g.*, diagnoses, medications) but also patient-specific context, such as their communication preferences, emotional state, and key concerns expressed over time; and (ii) Integrating patient-provider dialogue: With patient consent, the LLM could be aware of key advice from clinicians. This would enable it to act as a more effective translator and scaffolder, helping patients turn "the clinician's advice into specific daily actions" without contradicting the core treatment plan. This reinforces the LLM's complementary role and helps maintain a cohesive care narrative [32, 76]. Crucially, grounding the LLM's continuous support in the clinician's expert guidance serves as a critical check against the systemic risk of the LLM subtly misdirecting a patient's journey over time.

Designing for Responsible Empowerment and Trust Calibration. Our concept of fragile agency highlights a critical ethical tension: LLMs empower patients but also expose them to risks from misinformation, as they may lack the expertise to vet the LLM's advice. The design goal must shift from simply warning users about inaccuracies to actively designing for responsible empowerment. This is especially vital given the risks are not just overt error but also misplaced calm that can delay care or subtle misdirection that can flaw communication with clinicians. Existing work on human-AI collaboration suggests that explicitly signaling system limitations or uncertainty can effectively calibrate user trust

and reduce automation bias [78, 105, 158]. This requires moving beyond static disclaimers and creating interfaces that help patients calibrate their trust in real time. For instance, instead of delivering all information with the same confident tone, LLMs could be designed to express uncertainty explicitly [144]. When discussing less-established treatments or ambiguous symptoms, the UI could visually differentiate speculative information from evidence-based guidelines, with phrases such as "Here are a few possibilities being explored by researchers, but they are not standard practice...", acting as a direct countermeasure to the risk of false reassurance with overly confident tone [105]. Besides, rather than being an answer machine, the LLM could act as a Socratic partner. When a patient considers a significant action (*e.g.*, challenging a doctor's plan), the LLM could prompt critical reflection: "What are the main reasons you're considering this alternative? What are your biggest concerns about the doctor's original recommendation?" This approach empowers patients not by giving them answers, but by equipping them with better questions and a more structured thought process, thereby making their agency more robust [69, 147].

From Information-Giving to Scaffolding Health Literacy. Our findings show that one of the LLM's cognitive roles was to decompress dense medical advice into understandable, actionable knowledge (Section 4.4.1). Future designs should formalize this educational role by focusing on scaffolding long-term health literacy, as emphasized by the personal informatics community on going beyond providing data but supporting users' sensemaking [95, 109]. To ensure genuine understanding, the LLM could employ a "teach-back" method, a proven clinical communication technique [120, 145]. After explaining a concept, it could ask the patient to tell the LLM back in the patient's own words what this means. This transforms passive information consumption into active learning. Such interactive validation is a crucial safeguard, ensuring the patient has truly understood the nuances of the advice, and mitigating the risk of them acting on subtly misinterpreted or incorrectly "decompressed" guidance. Meanwhile, LLM systems could move from providing lists of suggestions to collaboratively creating personalized plans with the patient. For example, after discussing post-operative recovery, the LLM could help the patient co-author a simple daily schedule that incorporates the doctor's advice on medication, diet, and light exercise, making abstract guidance concrete and integrated into their lived reality. In this way, the LLM serves as a sustained partner, building the patient's capacity and confidence to navigate their own health journeys.

Self-Guardian Infrastructures to Improve Reliability. In addition to the patient-facing aspects of design, adopting a self-guardian and self-tracking infrastructure is a potential solution to reduce risks from the technical perspective. While the above suggestions focus on shaping the human-AI interaction, the internal architecture of the LLM system itself can be re-envisioned to be inherently safer and more reliable, reflecting the broader "Human-Centered AI" framework, which advocates for reliable, auditable, and safe system architectures as a prerequisite for user trust [5, 121]. A multi-agent system is a viable solution for this [58, 74], moving beyond a single LLM to a collaborative ensemble of specialized agents that check and balance one another, where distinct functions would be responsible for oversight. Some capabilities of such an infrastructure would include: continuous internal validation (systematically cross-referencing generated content against authenticated medical knowledge bases or established clinical guidelines), and dynamic safety monitoring (assessing conversational context to identify potential risks). By embedding these self-monitoring and corrective functions directly into the system's architecture, the burden of validation is shifted away from the patient. This approach provides a more robust technical foundation for responsible empowerment, allowing the LLM to more safely fulfill its complex roles as a longitudinal healthcare companion.

5.6 Limitations and Future Work

Our study has several limitations that point to important directions for future work. First, our qualitative analysis is based on patients' diary reflections. While the data is rich, they do not include a fine-grained analysis of the verbatim patient-LLM conversation data that could reveal deeper interactional nuances. Second, as discussed above, our study was conducted with 25 patients in China using specific LLMs available at the time; findings may differ across other cultural contexts, healthcare systems, and with more advanced or specialized LLM technologies. We posit that our core findings are likely generalizable to other healthcare systems facing similar pressures of time-scarcity and provider burnout globally [103, 140], yet future cross-cultural research is needed to disentangle these universal technological affordances from culturally specific adoption patterns. Meanwhile, participants who joined the study may inherently favor the adoption of LLMs, which may lead to potential bias in our data analysis. Additionally, while our study spanned four weeks (which is already longer or comparable to other similar studies [46, 60]), future longer-term research could uncover the evolution over an extended timeframe of patients' sense of connection with the LLM, perceived advice helpfulness, and reliance patterns. Furthermore, our findings are centered entirely on the patient's perspective, intentionally bracketing the crucial viewpoint of clinicians. Future research should therefore analyze these dynamics from multiple perspectives, including those of healthcare providers, to develop a holistic understanding of the reconfigured patient–doctor–AI triad.

6 Conclusion

In this paper, we presented a longitudinal diary study to examine how patients integrate LLMs into their healthcare-seeking journeys. Our findings demonstrated that patients adopt LLMs not as simple diagnostic or decision-support tools, but as dynamic longitudinal boundary companions that scaffold their experiences across informational, behavioral, emotional, and cognitive levels. This integration empowers patients as more active negotiators of their own care by reshaping relationships of agency, trust, and power. Our work suggests future directions for creating socio-technical systems that support relational continuity, foster responsible empowerment, and contribute to a safer and collaborative healthcare ecosystem across patients, LLMs, and clinicians.

Acknowledgments

Research reported in this publication was supported in part by the National Science Foundation (NSF) SCH under Award Number 2306690. The content is solely the responsibility of the authors and

does not necessarily represent the official views of the National Science Foundation. We also thank the Columbia Research Stabilization Grant for supporting this research project. We want to thank all participants for taking part in our study. We acknowledge the use of Generative AI tools to improve the grammar, style, and readability of this manuscript. These tools were used exclusively for text editing and played no role in the qualitative data analysis, interpretation, or generation of the core findings presented.

References

[1] 2025. https://www.yibiaoda.com/privacy
[2] Muhammad Aurangzeb Ahmad, Ilker Yaramis, and Taposh Dutta Roy. 2023. Creating trustworthy llms: Dealing with hallucinations in healthcare ai. *arXiv preprint arXiv:2311.01463* (2023).
[3] Chiwon Ahn. 2023. Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation* 185 (2023).
[4] Mutahar Ali, Arjun Arunasalam, and Habiba Farrukh. 2025. Understanding Users' Security and Privacy Concerns and Attitudes Towards Conversational AI Platforms. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 298–316.
[5] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
[6] Kanhai Amin, Pavan Khosla, Rushabh Doshi, Sophie Chheang, and Howard P Forman. 2023. Artificial intelligence to improve patient understanding of radiology reports. *The Yale journal of biology and medicine* 96, 3 (2023), 407.
[7] Ronald M Andersen. 1995. Revisiting the behavioral model and access to medical care: does it matter? *Journal of health and social behavior* (1995), 1–10.
[8] Antonis A Armoundas and Joseph Loscalzo. 2025. Patient agency and large language models in worldwide encoding of equity. *npj Digital Medicine* 8, 1 (2025), 258.
[9] Margaret Arnd-Caddigan. 2015. Sherry Turkle: Alone Together: Why We Expect More from Technology and Less from Each Other: Basic Books, New York, 2011, 348 pp, ISBN 978-0465031467 (pbk).
[10] Serhat Aydin, Mert Karabacak, Victoria Vlachos, and Konstantinos Margetis. 2024. Large language models in patient education: a scoping review of applications in medicine. *Frontiers in medicine* 11 (2024), 1477898.
[11] Serhat Aydin, Mert Karabacak, Victoria Vlachos, and Konstantinos Margetis. 2025. Navigating the potential and pitfalls of large language models in patient-centered medication guidance and self-decision support. *Frontiers in Medicine* 12 (2025), 1527864.
[12] Michael Balint. 1955. The doctor, his patient, and the illness. *The lancet* 265, 6866 (1955), 683–688.
[13] Elmer V Bernstam, Dawn M Shelton, Muhammad Walji, and Funda Meric-Bernstam. 2005. Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? *International journal of medical informatics* 74, 1 (2005), 13–19.
[14] Karthik S Bhat, Amanda K Hall, Tiffany Kuo, and Neha Kumar. 2023. " We are half-doctors": family caregivers as boundary actors in chronic disease management. *Proceedings of the ACM on human-computer interaction* 7, CSCW1 (2023), 1–29.
[15] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2010. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence* 24, 6 (2010), 648–666.
[16] Timothy W Bickmore, Lisa Caruso, Kerri Clough-Gorr, and Tim Heeren. 2005. 'It's just like you talk to a friend'relational agents for older adults. *Interacting with Computers* 17, 6 (2005), 711–735.
[17] Timothy W Bickmore, Rukmal Fernando, Lazlo Ring, and Daniel Schulman. 2010. Empathic touch by relational agents. *IEEE Transactions on Affective Computing* 1, 1 (2010), 60–71.
[18] Timothy W Bickmore and Rosalind W Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 293–327.
[19] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. 2022. My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research* 48, 3 (2022), 404–429.
[20] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712* (2023).
[21] Michael Bui, Kira Oberschmidt, and Christiane Grünloh. 2023. Patient Journey Value Mapping: Illustrating values and experiences along the patient journey to support eHealth design. In *Proceedings of Mensch und Computer 2023*. 49–66.
[22] Felix Busch, Lena Hoffmann, Christopher Rueger, Elon HC van Dijk, Rawen Kader, Esteban Ortiz-Prado, Marcus R Makowski, Luca Saba, Martin Hadamitzky, Jakob Nikolas Kather, et al. 2025. Current applications and challenges in large language models for patient care: a systematic review. *Communications Medicine* 5, 1 (2025), 26.
[23] Nicolas Carl, Sarah Haggenmüller, Christoph Wies, Lisa Nguyen, Jana Theres Winterstein, Martin Joachim Hetz, Maurin Helen Mangold, Friedrich Otto Hartung, Britta Grüne, Tim Holland-Letz, et al. 2025. Evaluating interactions of patients with large language models for medical information. *BJU international* 135, 6 (2025), 1010–1017.
[24] Marco Cascella, Jonathan Montomoli, Valentina Bellini, and Elena Bignami. 2023. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems* 47, 1 (March 2023), 33. https://doi.org/10.1007/s10916-023-01925-4
[25] David Chen, Kabir Chauhan, Rod Parsa, Zhihui Amy Liu, Fei-Fei Liu, Ernie Mak, Lawson Eng, Breffni Louise Hannon, Jennifer Croke, Andrew Hope, et al. 2025. Patient perceptions of empathy in physician and artificial intelligence chatbot responses to patient questions about cancer. *npj Digital Medicine* 8, 1 (2025), 275.
[26] Jiang Chen, Zhuochen Lin, Li-an Li, Jing Li, Yuyao Wang, Yu Pan, Jie Yang, Chuncong Xu, Xiaojing Zeng, Xiaoxu Xie, et al. 2021. Ten years of China's new healthcare reform: a longitudinal study on changes in health resources. *BMC public health* 21, 1 (2021), 2272.
[27] Jiaju Chen, Minglong Tang, Yuxuan Lu, Bingsheng Yao, Elissa Fan, Xiaojuan Ma, Ying Xu, Dakuo Wang, Yuling Sun, and Liang He. 2025. Characterizing LLM-Empowered Personalized Story Reading and Interaction for Children: Insights From Multi-Stakeholder Perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–24.
[28] Yu Chen, Tingxin Li, Huiming Liu, and Yang Yu. 2023. Hide and seek (has): A lightweight framework for prompt privacy protection. *arXiv preprint arXiv:2309.03057* (2023).
[29] Yuchen Chen, Silvia Lindtner, and Yuling Sun. 2025. The Moral Economy of AI. In *Proceedings of the sixth decennial Aarhus conference: Computing X Crisis*. 117–126.
[30] Yuchen Chen, Yuling Sun, and Silvia Lindtner. 2023. Maintainers of stability: The labor of China's data-driven governance and dynamic zero-COVID. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
[31] Minh Duc Chu, Patrick Gerard, Kshitij Pawar, Charles Bickham, and Kristina Lerman. 2025. Illusions of intimacy: Emotional attachment and emerging psychological risks in human-ai relationships. *arXiv preprint arXiv:2505.11649* (2025).
[32] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A Kientz, and Sean A Munson. 2016. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 770–786.
[33] Jan Clusmann, Fiona R Kolbinger, Hannah Sophie Muti, Zunamys I Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P Veldhuizen, et al. 2023. The future landscape of large language models in medicine. *Communications medicine* 3, 1 (2023), 141.
[34] Mary L Cummings. 2017. Automation bias in intelligent time critical decision support systems. In *Decision making in aviation*. Routledge, 289–294.
[35] Ellen L Davies, Lemma N Bulto, Alison Walsh, Danielle Pollock, Vikki M Langton, Robert E Laing, Amy Graham, Melissa Arnold-Chamney, and Janet Kelly. 2023. Reporting and conducting patient journey mapping research in healthcare: A scoping review. *Journal of advanced nursing* 79, 1 (2023), 83–100.
[36] Marnie de Mooij, Olivia Foss, and Brian Brost. 2022. Integrating the experience: Principles for digital transformation across the patient journey. *Digital Health* 8 (2022), 20552076221089100.
[37] Henrik HJ Detjen, Lars Densky, Niklas von Kalckreuth, and Marvin Kopka. 2025. Who is Trusted for a Second Opinion? Comparing Collective Advice from a Medical AI and Physicians in Biopsy Decisions After Mammography Screening. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–15.
[38] Roshini Deva, Dhruv Ramani, Tanvi Divate, Suhani Jalota, and Azra Ismail. 2025. "Kya family planning after marriage hoti hai?": Integrating Cultural Sensitivity in an LLM Chatbot for Reproductive Health. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.
[39] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O Riedl. 2022. Human-Centered Explainable AI (HCXAI): beyond opening the black-box of AI. In *CHI conference on human factors in computing systems extended abstracts*. 1–7.
[40] Madeleine Clare Elish, Saiph Savage, and Abeba Birhane. 2021. The hidden work created by artificial intelligence programs. *MIT Sloan Management Review: Ideas Made to Matter* (04 05 2021). https://mitsloan.mit.edu/ideas-made-to-matter/hidden-work-created-

artificial-intelligence-programs?utm_source=chatgpt.com Focuses on overlooked human labor in AI implementation, including "repair work" for workplace integration, "ghost workers" in behind-the-scenes tasks, and ethical considerations for marginalized groups affected by AI systems.

[41] Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130* (2023).

[42] Glyn Elwyn, Dominick Frosch, Richard Thomson, Natalie Joseph-Williams, Amy Lloyd, Paul Kinnersley, Emma Cording, Dave Tomson, Carole Dodd, Stephen Rollnick, et al. 2012. Shared decision making: a model for clinical practice. *Journal of general internal medicine* 27, 10 (2012), 1361–1367.

[43] Pouyan Esmaeilzadeh, Mahed Maddah, and Tala Mirzaei. 2025. Using AI chatbots (eg, CHATGPT) in seeking health-related information online: The case of a common ailment. *Computers in Human Behavior: Artificial Humans* 3 (2025), 100127.

[44] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.

[45] Charisse Foo, Pin Sym Foong, Camille Nadal, Natasha Ureyang, Thant Naylin, and Gerald Choon Huat Koh. 2025. The Benefits and Risks of LLMs for Facilitating Medical Decision-Making Among Laypersons. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 3173–3191.

[46] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. From text to self: users' perception of AIMC tools on interpersonal communication and self. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[47] Russell E Glasgow, Amy G Huebschmann, Alex H Krist, and Frank V Degruy. 2019. An adaptive, contextual, technology-aided support (ACTS) system for chronic illness self-management. *The Milbank Quarterly* 97, 3 (2019), 669–691.

[48] Ethan Goh, Robert J Gallo, Eric Strong, Yingjie Weng, Hannah Kerman, Jason A Freed, Joséphine A Cool, Zahir Kanjee, Kathleen P Lane, Andrew S Parsons, et al. 2025. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. *Nature Medicine* (2025), 1–6.

[49] Government Digital Service. [n. d.]. AI Insights: Prompt Engineering (HTML). https://www.gov.uk/government/publications/ai-insights/ai-insights-prompt-engineering-html Last updated on 5 June 2025; accessed on 2 September 2025.

[50] Edward Guadagnoli and Patricia Ward. 1998. Patient participation in decision-making. *Social science & medicine* 47, 3 (1998), 329–339.

[51] Xinning Gui and Yunan Chen. 2019. Making healthcare infrastructure work: Unpacking the infrastructuring work of individuals. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[52] Youqi Guo, Shu Hu, and Fei Liang. 2021. The prevalence and stressors of job burnout among medical staff in Liaoning, China: a cross-section study. *BMC public health* 21, 1 (2021), 777.

[53] Jennifer Haase and Paul HP Hanel. 2023. Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity* 33, 3 (2023), 100066.

[54] Joschka Haltaufderheide and Robert Ranisch. 2024. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). *NPJ digital medicine* 7, 1 (2024), 183.

[55] Gaole He, Lucie Kuiper, and Ujwal Gadiraju. 2023. Knowing about knowing: An illusion of human competence can hinder appropriate reliance on AI systems. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–18.

[56] Qian He, Fei Du, Lianne WL Simonse, et al. 2021. A patient journey map to improve the home isolation experience of persons with mild COVID-19: design research for service touchpoints of artificial intelligence in eHealth. *JMIR medical informatics* 9, 4 (2021), e23238.

[57] John Heritage and Douglas W Maynard. 2006. *Communication in medical care: Interaction between primary care physicians and patients*. Vol. 20. Cambridge University Press.

[58] A Ali Heydari, Ken Gu, Vidya Srinivas, Hong Yu, Zhihan Zhang, Yuwei Zhang, Akshay Paruchuri, Qian He, Hamid Palangi, Nova Hammerquist, et al. 2025. The Anatomy of a Personal Health Agent. *arXiv preprint arXiv:2508.20148* (2025).

[59] Tomasz Hollanek and Aisha Sobey. 2025. AI Companions for Health and Mental Wellbeing: Opportunities, Risks and Policy Implications. (2025).

[60] Matthew K Hong, Udaya Lakshmi, Kimberly Do, Sampath Prahalad, Thomas Olson, Rosa I Arriaga, and Lauren Wilcox. 2020. Using diaries to probe the illness experiences of adolescent patients and parental caregivers. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–16.

[61] Jieman Hu, Xue Wang, Shaoning Guo, Fangfang Chen, Yuan-yu Wu, Fu-jian Ji, and Xuedong Fang. 2019. Peer support interventions for breast cancer patients: a systematic review. *Breast cancer research and treatment* 174, 2 (2019), 325–341.

[62] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. 2022. Creative writing with an ai-powered writing assistant: Perspectives from professional writers. *arXiv preprint arXiv:2211.05030* (2022).

[63] Maia Jacobs, James Clawson, and Elizabeth D Mynatt. 2016. A cancer journey framework: guiding the design of holistic health technology. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 114–121.

[64] Katharina Jeblick, Balthasar Schachtner, Jakob Dexl, Andreas Mittermeier, Anna Theresa Stüber, Johanna Topalis, Tobias Weber, Philipp Wesp, Bastian Oliver Sabel, Jens Ricke, et al. 2024. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *European radiology* 34, 5 (2024), 2817–2825.

[65] Eunkyung Jo, Yuin Jeong, SoHyun Park, Daniel A Epstein, and Young-Ho Kim. 2024. Understanding the impact of long-term memory on self-disclosure with large language model-driven chatbots for public health intervention. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.

[66] Kyuha Jung, Gyuho Lee, Yuanhui Huang, and Yunan Chen. 2025. 'I've talked to ChatGPT about my issues last night.': Examining Mental Health Conversations with Large Language Models through Reddit Analysis. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–25.

[67] Riyaz Kaba and Prasanna Sooriakumaran. 2007. The evolution of the doctor-patient relationship. *International journal of surgery* 5, 1 (2007), 57–65.

[68] Zhigang Kan, Linbo Qiao, Hao Yu, Liwen Peng, Yifu Gao, and Dongsheng Li. 2023. Protecting user privacy in remote conversational systems: A privacy-preserving framework based on text sanitization. *arXiv preprint arXiv:2306.08223* (2023).

[69] Hyunjin Kang and Chen Lou. 2022. AI agency vs. human agency: understanding human–AI interactions on TikTok and their implications for user engagement. *Journal of Computer-Mediated Communication* 27, 5 (2022), zmac014.

[70] Elizabeth Kaziunas, Mark S Ackerman, Silvia Lindtner, and Joyce M Lee. 2017. Caring through data: Attending to the social and emotional experiences of health datafication. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2260–2272.

[71] Dajung Kim, Niko Vegt, Valentijn Visch, and Marina Bos-De Vos. 2024. How Much Decision Power Should (A) I Have?: Investigating Patients' Preferences Towards AI Autonomy in Healthcare Decision Making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[72] Jiyeong Kim, Stephen P Ma, Michael L Chen, Isaac R Galatzer-Levy, John Torous, Peter J van Roessel, Christopher Sharp, Michael A Pfeffer, Carolyn I Rodriguez, Eleni Linos, et al. 2025. Optimizing large language models for detecting symptoms of depression/anxiety in chronic diseases patient communications. *NPJ Digital Medicine* 8, 1 (2025), 580.

[73] Siwon Kim, Sangdoo Yun, Hwaran Lee, Martin Gubri, Sungroh Yoon, and Seong Joon Oh. 2023. Propile: Probing privacy leakage in large language models. *Advances in Neural Information Processing Systems* 36 (2023), 20750–20762.

[74] Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik S Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae W Park. 2024. Mdagents: An adaptive collaboration of llms for medical decision-making. *Advances in Neural Information Processing Systems* 37 (2024), 79410–79452.

[75] Yubin Kim, Xuhai Xu, Daniel McDuff, Cynthia Breazeal, and Hae Won Park. 2024. Health-llm: Large language models for health prediction via wearable sensor data. *arXiv preprint arXiv:2401.06866* (2024).

[76] Susanne Kirchner, Jessica Schroeder, James Fogarty, and Sean A Munson. 2021. "They don't always think about that": Translational Needs in the Design of Personal Health Informatics Applications. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–16.

[77] Rob Kling, Geoffrey McKim, Joanna Fortuna, and Adam King. 2000. Scientific collaboratories as socio-technical interaction networks: a theoretical approach. *arXiv preprint cs/0005007* (2000).

[78] Rafal Kocielnik, Saleema Amershi, and Paul N Bennett. 2019. Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[79] Ashutosh Kumar, Shiv Vignesh Murthy, Sagarika Singh, and Swathy Ragupathy. 2024. The ethics of interaction: Mitigating security threats in llms. *arXiv preprint arXiv:2401.12273* (2024).

[80] Jabari Kwesi, Jiaxun Cao, Riya Manchanda, and Pardis Emami-Naeini. 2025. Exploring user security and privacy attitudes and concerns toward the use of {General-Purpose}{LLM} chatbots for mental health. In *34th USENIX Security Symposium (USENIX Security 25)*. 6007–6024.

[81] Moustafa Laymouna, Yuanchao Ma, David Lessard, Tibor Schuster, Kim Engler, and Bertrand Lebouché. 2024. Roles, users, benefits, and limitations of chatbots in health care: rapid review. *Journal of medical Internet research* 26 (2024), e56930.

[82] Jamie Lee, Kyuha Jung, Erin Gregg Newman, Emilie Chow, and Yunan Chen. 2025. Understanding Adolescents' Perceptions of Benefits and Risks in Health AI Technologies through Design Fiction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[83] Ying Lei, Shuai Ma, Yuling Sun, and Xiaojuan Ma. 2025. "AI Afterlife" as Digital Legacy: Perceptions, Expectations, and Concerns. In *Proceedings of the 2025 CHI*

Conference on Human Factors in Computing Systems. 1–18.
[84] Susan Leigh Star. 2010. This is not a boundary object: Reflections on the origin of a concept. *Science, technology, & human values* 35, 5 (2010), 601–617.
[85] Caixia Li, Yina Zhao, Yang Bai, Baoquan Zhao, Yetunde Oluwafunmilayo Tola, Carmen WH Chan, Meifen Zhang, and Xia Fu. 2025. Unveiling the potential of large language models in transforming chronic disease management: mixed methods systematic review. *Journal of Medical Internet Research* 27 (2025), e70535.
[86] Jia Li, Zichun Zhou, Han Lyu, and Zhenchang Wang. 2025. Large language models-powered clinical decision support: enhancing or replacing human expertise? , 4 pages.
[87] Tianshi Li, Sauvik Das, Hao-Ping Lee, Dakuo Wang, Bingsheng Yao, and Zhiping Zhang. 2024. Human-centered privacy research in the age of large language models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–4.
[88] Huigang Liang, Yajiong Xue, and Zhi-ruo Zhang. 2021. Patient satisfaction in China: a national survey of inpatients and outpatients. *BMJ open* 11, 9 (2021), e049570.
[89] Wei Liu, Jun Li, Yitao Tang, Yining Zhao, Chaozhong Liu, Meiyi Song, Zhenlin Ju, Shwetha V Kumar, Yiling Lu, Rehan Akbani, et al. 2025. DrBioRight 2.0: an LLM-powered bioinformatics chatbot for large-scale cancer functional proteomics analysis. *Nature communications* 16, 1 (2025), 2256.
[90] Xiaohong Liu, Hao Liu, Guoxing Yang, Zeyu Jiang, Shuguang Cui, Zhaoze Zhang, Huan Wang, Liyuan Tao, Yongchang Sun, Zhu Song, et al. 2025. A generalist medical language model for disease diagnosis assistance. *Nature Medicine* (2025), 1–11.
[91] Zhihuang Liu, Ling Hu, Tongqing Zhou, Yonghao Tang, and Zhiping Cai. 2025. Prevalence Overshadows Concerns? Understanding Chinese Users' Privacy Awareness and Expectations Towards LLM-Based Healthcare Consultation. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2716–2734.
[92] Gang Luo, Chunqiang Tang, Hao Yang, and Xing Wei. 2008. MedSearch: a specialized search engine for medical information retrieval. In *Proceedings of the 17th ACM conference on Information and knowledge management*. 143–152.
[93] Xiaolei Lv, Xiaomeng Zhang, Yuan Li, Xinxin Ding, Hongchang Lai, and Junyu Shi. 2024. Leveraging large language models for improved patient access and self-management: assessor-blinded comparison between expert-and AI-generated content. *Journal of medical Internet research* 26 (2024), e55847.
[94] Xiaojuan Ma, Xinning Gui, Jiayue Fan, Mingqian Zhao, Yunan Chen, and Kai Zheng. 2018. Professional Medical Advice at your Fingertips: An empirical study of an online" Ask the Doctor" platform. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–22.
[95] Lena Mamykina, Arlene M Smaldone, and Suzanne R Bakken. 2015. Adopting the sensemaking perspective for chronic disease self-management. *Journal of biomedical informatics* 56 (2015), 406–417.
[96] Ashley ND Meyer, Traber D Giardina, Christiane Spitzmueller, Umber Shahid, Taylor MT Scott, and Hardeep Singh. 2020. Patient perspectives on the usefulness of an artificial intelligence–assisted symptom checker: cross-sectional survey study. *Journal of medical Internet research* 22, 1 (2020), e14679.
[97] Jonah Meyerhoff, Rachel Kornfield, David C Mohr, and Madhu Reddy. 2022. Meeting young adults' social support needs across the health behavior change journey: implications for digital mental health tools. *Proceedings of the ACM on Human-computer Interaction* 6, CSCW2 (2022), 1–33.
[98] Tala Mirzaei, Leila Amini, and Pouyan Esmaeilzadeh. 2024. Clinician voices on ethics of LLM integration in healthcare: a thematic analysis of ethical concerns and implications. *BMC Medical Informatics and Decision Making* 24, 1 (2024), 250.
[99] Rudolf H Moos and Vivien Davis Tsu. 1977. The crisis of physical illness: An overview. *Coping with physical illness* (1977), 3–21.
[100] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2016. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of behavioral medicine* (2016), 1–17.
[101] Zabir Al Nazi and Wei Peng. 2024. Large language models in healthcare and medical domain: A review. In *Informatics*, Vol. 11. MDPI, 57.
[102] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375* (2023).
[103] Writing Committee of the Report on Cardiovascular Health, Diseases in China, et al. 2022. Report on cardiovascular health and diseases in China 2021: an updated summary. *Biomedical and Environmental Sciences* 35, 7 (2022), 573–603.
[104] Yujin Oh, Sangjoon Park, Hwa Kyung Byun, Yeona Cho, Ik Jae Lee, Jin Sung Kim, and Jong Chul Ye. 2024. LLM-driven multimodal target volume contouring in radiation oncology. *Nature Communications* 15, 1 (2024), 9186.
[105] Kazuo Okamura and Seiji Yamada. 2020. Adaptive trust calibration for human-AI collaboration. *Plos one* 15, 2 (2020), e0229132.
[106] Oxford Insights. [n. d.]. LLMs in Government: Brainstorming Applications. https://oxfordinsights.com/insights/llms-in-government-brainstorming-applications/ Published 19 May 2023; accessed 2 September 2025.
[107] Talcott Parsons. 2013. *The social system*. Routledge.
[108] Rob Procter, Peter Tolmie, and Mark Rouncefield. 2023. Holding AI to account: challenges for the delivery of trustworthy AI in healthcare. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–34.
[109] Aare Puussaar, Adrian K Clear, and Peter Wright. 2017. Enhancing personal informatics through social sensemaking. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6936–6942.
[110] Jianing Qiu, Kyle Lam, Guohao Li, Amish Acharya, Tien Yin Wong, Ara Darzi, Wu Yuan, and Eric J Topol. 2024. LLM-based agentic systems in medicine and healthcare. *Nature Machine Intelligence* 6, 12 (2024), 1418–1420.
[111] Arya Rao, John Kim, Meghana Kamineni, Michael Pang, Winston Lie, and Marc D Succi. 2023. Evaluating ChatGPT as an adjunct for radiologic decision-making. *MedRxiv* (2023), 2023–02.
[112] Vishal Rathod, Seyedsina Nabavirazavi, Samira Zad, and Sundararaja Sitharama Iyengar. 2025. Privacy and security challenges in large language models. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 00746–00752.
[113] Mubashar Raza, Zarmina Jahangir, Muhammad Bilal Riaz, Muhammad Jasim Saeed, and Muhammad Awais Sattar. 2025. Industrial applications of large language models. *Scientific Reports* 15, 1 (2025), 13755.
[114] Laurel D Riek and Lilly Irani. 2025. The Future Is Rosie?: Disempowering Arguments About Automation and What to Do About It. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–14.
[115] Nudtaporn Rosruen and Taweesak Samanchuen. 2018. Chatbot utilization for medical consultant system. In *2018 3rd technology innovation management and engineering science international conference (TIMES-iCON)*. IEEE, 1–5.
[116] Malik Sallam. 2023. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare*, Vol. 11. MDPI, 887.
[117] Ashish Sarraju, Dennis Bruemmer, Erik Van Iterson, Leslie Cho, Fatima Rodriguez, and Luke Laffin. 2023. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *Jama* 329, 10 (2023), 842–844.
[118] SE Scott, FM Walter, Andrew Webster, Stephen Sutton, and Jon Emery. 2013. The model of pathways to treatment: conceptualization and integration with existing theory. *British journal of health psychology* 18, 1 (2013), 45–65.
[119] Hannah L Semigran, Jeffrey A Linder, Courtney Gidengil, and Ateev Mehrotra. 2015. Evaluation of symptom checkers for self diagnosis and triage: audit study. *bmj* 351 (2015).
[120] Violetta Shersher, Terry P Haines, Liz Sturgiss, Carolina Weller, and Cylie Williams. 2021. Definitions and use of the teach-back method in healthcare consultations with patients: A systematic review and thematic synthesis. *Patient education and counseling* 104, 1 (2021), 118–129.
[121] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.
[122] Sina Shool, Sara Adimi, Reza Saboori Amleshi, Ehsan Bitaraf, Reza Golpira, and Mahmood Tara. 2025. A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Medical Informatics and Decision Making* 25, 1 (2025), 117.
[123] Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. 2023. Large language models encode clinical knowledge. *Nature* 620, 7972 (2023), 172–180.
[124] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R Pfohl, Heather Cole-Lewis, et al. 2025. Toward expert-level medical question answering with large language models. *Nature Medicine* (2025), 1–8.
[125] Ranwir K Sinha, Asitava Deb Roy, Nikhil Kumar, Himel Mondal, and Ranwir Sinha. 2023. Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus* 15, 2 (2023).
[126] Bryan A Sisk, Christine Bereitschaft, Stephanie Milam, Anna M Kerr, Justin N Baker, Theodore Marghitu, Beatrice White, Jennifer W Mack, Fabienne Bourgeois, and James M DuBois. 2026. Trust, Transparency, and Acceptability of Artificial Intelligence-Powered Communication Tools in Cancer: Focus Groups With Parents, Adolescents, and Young Adults. *Pediatric Blood & Cancer* 73, 1 (2026), e32094.
[127] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies* 149 (2021), 102601.
[128] Inhwa Song, Sachin R Pendse, Neha Kumar, and Munmun De Choudhury. 2025. The typing cure: Experiences with large language model chatbots for mental health support. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–29.
[129] Catherine J Staes, Anna C Beck, George Chalkidis, Carolyn H Scheese, Teresa Taft, Jia-Wen Guo, Michael G Newman, Kensaku Kawamoto, Elizabeth A Sloss, and Jordan P McPherson. 2024. Design of an interface to communicate artificial

intelligence-based prognosis for patients with advanced solid tumors: a user-centered approach. *Journal of the American Medical Informatics Association* 31, 1 (2024), 174–187.

[130] Susan Leigh Star and Karen Ruhleder. 1994. Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. 253–264.

[131] Charumathi Raghu Subramanian, Daniel A Yang, and Raman Khanna. 2024. Enhancing health care communication with large language models—the role, challenges, and future directions. *JAMA network open* 7, 3 (2024), e240347–e240347.

[132] Lucille Alice Suchman. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.

[133] Lucille Alice Suchman. 2007. *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.

[134] Jina Suh, Spencer Williams, Jesse R Fann, James Fogarty, Amy M Bauer, and Gary Hsieh. 2020. Parallel journeys of patients with cancer and depression: challenges and opportunities for technology-enabled collaborative care. *Proceedings of the ACM on Human-computer Interaction* 4, CSCW1 (2020), 1–36.

[135] Yuling Sun, Sam Addison Ankenbauer, Zhifan Guo, Yuchen Chen, Xiaojuan Ma, and Liang He. 2025. Rethinking Technological Solutions for Community-Based Older Adult Care: Insights fromOlder Partners' in China. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–36.

[136] 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.

[137] Yuling Sun, Xiaojuan Ma, Silvia Lindtner, and Liang He. 2023. Care Workers' Wellbeing in Data-Driven Healthcare Workplace: Identity, Agency, and Social Justice. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–29.

[138] 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.

[139] 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. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–37.

[140] Ming Tai-Seale, Thomas G McGuire, and Weimin Zhang. 2007. Time allocation in primary care office visits. *Health services research* 42, 5 (2007), 1871–1894.

[141] Tara Templin, Monika W Perez, Sean Sylvia, Jeff Leek, and Nasa Sinnott-Armstrong. 2024. Addressing 6 challenges in generative AI for digital health: a scoping review. *PLOS digital health* 3, 5 (2024), e0000503.

[142] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine* 29, 8 (2023), 1930–1940.

[143] Constantinos Timinis, Jeremy Opie, Simon Watt, Yvonne Rogers, and Ivana Drobnjak. 2025. Co-Creating Reassurance Journey Maps to Foster Engagement in Remote Patient Monitoring for Post-Operative Cancer Care. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.

[144] Richard Tomsett, Alun Preece, Dave Braines, Federico Cerutti, Supriyo Chakraborty, Mani Srivastava, Gavin Pearson, and Lance Kaplan. 2020. Rapid trust calibration through interpretable and uncertainty-aware AI. *Patterns* 1, 4 (2020).

[145] Sophie Tran, Gabrielle Bennett, Jacqui Richmond, Tin Nguyen, Marno Ryan, Thai Hong, Jessica Howell, Barbara Demediuk, Paul Desmond, Sally Bell, et al. 2019. 'Teach-back'is a simple communication tool that improves disease knowledge in people with chronic hepatitis B–a pilot randomized controlled study. *BMC Public Health* 19, 1 (2019), 1355.

[146] Mara Ulloa, Blaine Rothrock, Faraz S Ahmad, and Maia Jacobs. 2022. Invisible clinical labor driving the successful integration of AI in healthcare. *Frontiers in Computer Science* 4 (2022), 1045704.

[147] Usman Ahmad Usmani, Ari Happonen, and Junzo Watada. 2023. Human-centered artificial intelligence: Designing for user empowerment and ethical considerations. In *2023 5th international congress on human-computer interaction, optimization and robotic applications (HORA)*. IEEE.

[148] Sterre Van Arum, Hüseyin Uğur Genç, Dennis Reidsma, and Armağan Karahanoğlu. 2025. Selective Trust: Understanding Human-AI Partnerships in Personal Health Decision-Making Process. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.

[149] Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Christian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerová, et al. 2024. Adapted large language models can outperform medical experts in clinical text summarization. *Nature medicine* 30, 4 (2024), 1134–1142.

[150] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S Bernstein, and Ranjay Krishna. 2023. Explanations can reduce overreliance on ai systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–38.

[151] Nervo Verdezoto, Naveen Bagalkot, Syeda Zainab Akbar, Swati Sharma, Nicola Mackintosh, Deirdre Harrington, and Paula Griffiths. 2021. The invisible work of maintenance in community health: challenges and opportunities for digital health to support frontline health workers in Karnataka, South India. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.

[152] Josip Vrdoljak, Zvonimir Boban, Ivan Males, Roko Skrabic, Marko Kumric, Anna Ottosen, Alexander Clemencau, Josko Bozic, and Sebastian Völker. 2025. Evaluating large language and large reasoning models as decision support tools in emergency internal medicine. *Computers in Biology and Medicine* 192 (2025), 110351.

[153] Boxin Wang, Weixin Chen, Hengzhi Pei, Chulin Xie, Mintong Kang, Chenhui Zhang, Chejian Xu, Zidi Xiong, Ritik Dutta, Rylan Schaeffer, et al. 2023. DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models. In *NeurIPS*.

[154] Bingcheng Wang, Pei-Luen Patrick Rau, and Tianyi Yuan. 2023. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology* 42, 9 (2023), 1324–1337.

[155] 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.

[156] Haochun Wang, Sendong Zhao, Zewen Qiang, Nuwa Xi, Bing Qin, and Ting Liu. 2024. Beyond direct diagnosis: LLM-based multi-specialist agent consultation for automatic diagnosis. *arXiv preprint arXiv:2401.16107* (2024).

[157] Jinge Wang, Kenneth Shue, Li Liu, and Gangqing Hu. 2025. Preliminary evaluation of ChatGPT model iterations in emergency department diagnostics. *Scientific reports* 15, 1 (2025), 10426.

[158] Lifei Wang, Natalie Friedman, Chengchao Zhu, Zeshu Zhu, and S Joy Mountford. 2025. The Impact of Confidence Ratings on User Trust in Large Language Models. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 365–370.

[159] Isabella C Wiest, Marie-Elisabeth Leßmann, Fabian Wolf, Dyke Ferber, Marko Van Treeck, Jiefu Zhu, Matthias P Ebert, Christoph Benedikt Westphalen, Martin Wermke, and Jakob Nikolas Kather. 2025. Deidentifying medical documents with local, privacy-preserving large language models: the LLM-anonymizer. *NEJM AI* 2, 4 (2025), AIdbp2400537.

[160] Wikipedia contributors. [n. d.]. Large language model. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Large_language_model Last edited on *May 26, 2025*; accessed on 2025-09-02.

[161] Lauren Wilcox, Renee Shelby, Rajesh Veeraraghavan, Oliver L Haimson, Gabriela Cruz Erickson, Michael Turken, and Rebecca Gulotta. 2023. Infrastructuring care: How trans and non-binary people meet health and well-being needs through technology. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[162] Ruolan Wu, Chun Yu, Xiaole Pan, Yujia Liu, Ningning Zhang, Yue Fu, Yuhan Wang, Zhi Zheng, Li Chen, Qiaolei Jiang, et al. 2024. MindShift: leveraging large language models for mental-states-based problematic smartphone use intervention. In *Proceedings of the 2024 CHI conference on human factors in computing systems*. 1–24.

[163] He Xiao, Xingjiao Wu, Jialiang Tong, Bangyan Li, and Yuling Sun. 2024. Chinese Elderly Healthcare-Oriented Conversation: CareQA Dataset and Its Knowledge Distillation Based Generation Framework. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 3866–3871.

[164] Yunpeng Xiao, Kyrie Zhixuan Zhou, Yueqing Liang, and Kai Shu. 2024. Understanding the concerns and choices of public when using large language models for healthcare. *arXiv preprint arXiv:2401.09090* (2024).

[165] Lu Xu, Leslie Sanders, Kay Li, and James CL Chow. 2021. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR cancer* 7, 4 (2021), e27850.

[166] Xuhai Xu, Bingsheng Yao, Yuanzhe Dong, Saadia Gabriel, Hong Yu, James Hendler, Marzyeh Ghassemi, Anind K Dey, and Dakuo Wang. 2024. Mental-llm: Leveraging large language models for mental health prediction via online text data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 1 (2024), 1–32.

[167] Jingjing Yan, Jiansen Yao, and Dahai Zhao. 2021. Patient satisfaction with outpatient care in China: a comparison of public secondary and tertiary hospitals. *International Journal for Quality in Health Care* 33, 1 (2021), mzab003.

[168] Siyuan Yan, Zhen Yu, Clare Primiero, Cristina Vico-Alonso, Zhonghua Wang, Litao Yang, Philipp Tschandl, Ming Hu, Lie Ju, Gin Tan, et al. 2025. A multimodal vision foundation model for clinical dermatology. *Nature Medicine* (2025), 1–12.

[169] Bufang Yang, Siyang Jiang, Lilin Xu, Kaiwei Liu, Hai Li, Guoliang Xing, Hongkai Chen, Xiaofan Jiang, and Zhenyu Yan. 2024. Drhouse: An llm-empowered

diagnostic reasoning system through harnessing outcomes from sensor data and expert knowledge. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 4 (2024), 1–29.

[170] Xintian Yang, Tongxin Li, Qin Su, Yaling Liu, Chenxi Kang, Yong Lyu, Lina Zhao, Yongzhan Nie, and Yanglin Pan. 2025. Application of large language models in disease diagnosis and treatment. *Chinese Medical Journal* 138, 02 (2025), 130–142.

[171] 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.

[172] Ruoyu Yin, Lishi Yin, Lin Li, Jennifer Silva-Nash, Jingru Tan, Zixian Pan, Jianying Zeng, and Lijing L Yan. 2022. Hypertension in China: burdens, guidelines and policy responses: a state-of-the-art review. *Journal of human hypertension* 36, 2 (2022), 126–134.

[173] Dong Whi Yoo, Jiayue Melissa Shi, Violeta J Rodriguez, and Koustuv Saha. 2025. AI Chatbots for Mental Health: Values and Harms from Lived Experiences of Depression. *arXiv preprint arXiv:2504.18932* (2025).

[174] Travis Zack, Eric Lehman, Mirac Suzgun, Jorge A Rodriguez, Leo Anthony Celi, Judy Gichoya, Dan Jurafsky, Peter Szolovits, David W Bates, Raja-Elie E Abdulnour, et al. 2023. Coding inequity: assessing GPT-4's potential for perpetuating racial and gender biases in healthcare. *MedRxiv* (2023), 2023–07.

[175] Santiago Zanella-Béguelin, Lukas Wutschitz, Shruti Tople, Victor Rühle, Andrew Paverd, Olga Ohrimenko, Boris Köpf, and Marc Brockschmidt. 2020. Analyzing information leakage of updates to natural language models. In *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 363–375.

[176] Yutong Zhang, Dora Zhao, Jeffrey T Hancock, Robert Kraut, and Diyi Yang. 2025. The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. *arXiv preprint arXiv:2506.12605* (2025).

[177] Zhiping Zhang, Michelle Jia, Hao-Ping Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–26.

[178] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* 1, 2 (2023).

[179] Qin Zheng, Kun Yang, Rui-Jie Zhao, Xue Wang, Ping Ping, Zheng-Hang Ou, Xiao-Peng Su, Jing Zhang, and Miao Qu. 2022. Burnout among doctors in China through 2020: A systematic review and meta-analysis. *Heliyon* 8, 7 (2022).

Appendix:

Table 3: Details of all LLMs used by participants in this study, categorized as general-purpose and health-focused.

LLM Category	Specific LLM Name	Developer	URL
General-purpose LLM	Doubao	ByteDance	https://www.doubao.com/chat/
	DeepSeek	DeepSeek AI	https://www.deepseek.com/
	Qwen	Alibaba Cloud	https://www.tongyi.com/
	Kimi	Moonshot AI	https://www.kimi.com/
	ChatGPT	OpenAI	https://chatgpt.com/
Health-focused LLM	Xiaohe Health	ByteDance	https://www.tiktok.com/explore
	iFlytek Xiaoyi	iFlytek	https://chatdr.iflyhealth.com/
	Baidu Lingyi Zhihui	Baidu	https://01.baidu.com/home/
	JD Health	JD Health	https://ir.jdhealth.com/sc/index.php
	Zuoshouyisheng	Zuoshouyisheng	https://zuoshouyisheng.com/
	AQ	Ant Group	https://www.alipay.com/

Table 4: Details of all diary questionnaires, including interaction contexts, the corresponding question categories, and the specific questions.

Interaction Context	Question Categories	Specific Question
Interaction with LLM(s)	Specifics	Did you consult more than one AI tool for health advice today?
		Which LLM(s) did you use?
	Rationale for the Interaction	What was your primary reason for seeking health advice from the LLM today?
		[If applicable] What were your reasons for consulting multiple LLMs?
	Conversation Content	Please provide the transcript (via link or screenshot) of your LLM conversation(s) today.
	Perceived Role	What role did the LLM play in your health trajectory?
	User Experience	Please describe your overall experience with the LLM interaction today.
		How did the LLM influence your emotional state today?
		Do you have any feedback on your interaction with the LLM today, such as unsatisfactory aspects or suggestions for future improvements?
Interaction with human doctor(s)	Reason for the Visit	What was your primary reason for consulting human doctor(s) today?
	Clinical Outcomes	Please summarize what you have discussed with the human doctor today.
	Perceived Role	What role did the human doctor's advice play in your health trajectory?
Interaction with both	Reason for Interacting with Both	What were your reasons for consulting both the LLM and the human doctor for your health concern today?
	Perceived Relationship between the Two	What were the key differences between the advice provided by the LLM and the human doctor?
		How do you perceive the relationship between the LLM and the human doctor in the context of the healthcare-seeking trajectory?
Others	-	Do you have any other things you would like to share?

Table 5: Participants' educational backgrounds and AI literacy levels. We report four representative items from the AI literacy scale, alongside the Total Score, which is calculated as the mean of all 12 items [154]. All scores in this table range from 1 to 7, with higher values indicating higher AI literacy.

ID	Education	Device Distinction	Tech Identification	Capability Evaluation	Misuse Vigilance	Total Score (averaged)
P01	Bachelor's	7/7	5/7	5/7	5/7	5.8/7
P02	Bachelor's	6/7	6/7	5/7	5/7	6.1/7
P03	Bachelor's	7/7	6/7	7/7	5/7	6.2/7
P04	Bachelor's	7/7	5/7	5/7	6/7	6.0/7
P05	Bachelor's	6/7	5/7	5/7	4/7	5.4/7
P06	Bachelor's	6/7	5/7	6/7	5/7	6.3/7
P07	Bachelor's	6/7	5/7	3/7	2/7	5.3/7
P08	Bachelor's	5/7	5/7	5/7	3/7	4.8/7
P09	Master's	7/7	7/7	7/7	7/7	6.6/7
P10	Bachelor's	5/7	6/7	6/7	5/7	5.9/7
P11	Bachelor's	7/7	6/7	6/7	4/7	6.5/7
P12	Bachelor's	6/7	7/7	6/7	3/7	5.1/7
P13	Bachelor's	6/7	5/7	6/7	5/7	6.0/7
P14	Bachelor's	7/7	7/7	7/7	1/7	6.5/7
P15	Bachelor's	6/7	7/7	7/7	2/7	6.0/7
P16	Bachelor's	7/7	5/7	5/7	1/7	5.3/7
P17	Bachelor's	7/7	5/7	4/7	5/7	6.0/7
P18	Bachelor's	7/7	6/7	5/7	5/7	5.8/7
P19	Bachelor's	7/7	5/7	5/7	7/7	6.5/7
P20	Bachelor's	6/7	6/7	5/7	5/7	5.4/7
P21	Bachelor's	7/7	3/7	7/7	4/7	5.8/7
P22	Bachelor's	7/7	7/7	7/7	2/7	5.9/7
P23	Bachelor's	6/7	6/7	4/7	5/7	5.7/7
P24	High School	4/7	5/7	2/7	1/7	5.3/7
P25	Bachelor's	5/7	6/7	6/7	6/7	5.9/7