Understanding use of Large Language Models in Healthcare: An HCI Scoping Review

Pradeepa Bandara

The University of Queensland, p.bandara@uq.edu.au

Thilina Halloluwa

The University of Queensland, t.halloluwa@uq.edu.au

Dhaval Vyas

The University of Queensland, d.vyas@uq.edu.au

Large Language Models (LLMs) and their integrations are transforming the healthcare sector, presenting both opportunities and challenges. Despite the increasing amount of research on improving the computational aspects of LLM algorithms, there is a significant gap in work focusing on the human element, especially how these technologies impact users and enhance their experiences. This paper provides a comprehensive scoping review of recent HCI research on the use of LLMs in healthcare. Our review synthesizes findings from papers sourced mainly from the ACM digital library highlighting human-centered implications of these technologies. Through a thematic analysis we have identified three key themes: supporting healthcare providers, enhancing patient interactions, and empowering communities through education. Our findings and future directions emphasize the importance of user-centered design in the integration of LLMs in healthcare. This paper serves as a foundational reference for future research within the HCI community, emphasizing the importance of developing LLMs that are not only accurate and efficient but also compassionate, ethical, responsible and user centered.

CCS CONCEPTS • Human-centered computing • Human Computer Interaction (HCI) • HCI Concepts, Theory and Models

Additional Keywords and Phrases: Large Language Models, HCI, Health

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

LLMs and Generative Artificial Intelligence (Gen AI) are revolutionizing various fields, including healthcare [10, 21, 26, 42] , by changing the way people work and interact with technology. They are capable of understanding and generating text and images, which enables professionals and relevant stakeholders to assist in a wide range of tasks. For example, within the health sector they are capable of tasks from diagnosing diseases [31, 46] to providing emotional support [33, 62]. They are even capable of interacting with multiple spoken languages (e.g. [15, 48]). As a result, there is an increasing amount of research on LLMs and various forms of Gen AI. Weizenbaum [70] delves into the impact of computers on human perception and society by exploring the development of conversational computer programs like Eliza and the potential for computerized psychotherapy, raising questions about the boundaries between human and machine intelligence. In another study by Kenneth Colby [9] focused on the development of a model of paranoid thinking within the realm of artificial intelligence research. However, even though there is substantial research on improving the computational efficiency of LLM algorithms (e.g. [13]) and even introducing new models (e.g. Open AI's GPT [84], Google's GEMINI [85], Meta's LlaMa [86]) there is a notable lack of work on how these technologies impact users and can enhance human lives. As Human Computer Interaction (HCI) practitioners, we believe that it is crucial to explore these developments from a human-centered perspective to ensure that the integration of LLMs prioritizes user needs, enhances user experiences, and addresses ethical considerations [67], especially when dealing with the health sector where many stakeholders are vulnerable.

The aim of this paper is to synthesize HCI research on the use of LLMs in healthcare to establish research directions for future research in HCI and to highlight the importance of keeping the user at the center of these technological advancements. We utilized the scoping review methodology [4] supplemented with the PRISMA model [64], as a research methodology to explore the existing literature on the topic, focusing on works published after January 2022. There are similar systematic reviews conducted on the use of LLMs in specific branches of the health sector and specific groups of users. For example, there are reviews conducted focusing on mental health [23, 47] , medicine [41], public health [27]. Additionally, Iloanusi and Chun [24] have explored the effects of LLMs on marginalized minorities. These literature surveys predominantly explored the computational aspects of the LLMs (e.g. [7, 41]) rather than the user perspectives and failed to capture papers published in important HCI venues, such as CHI and CSCW, among others. In contrast to these reviews, this paper provides a much broader and timely scoping review focusing specifically on the HCI literature. As far as we are aware, this is the first scoping review conducted on the use of LLMs in healthcare from a user-centered perspective.
Following are the research questions we explored with our scoping review.

RQ1*:* What is the discourse around the use of LLM powered tools and their impact on patient satisfaction, trust, and engagement in managing their health conditions?

RQ2*:* What are the experiences and challenges faced by healthcare providers when integrating LLMs into their clinical practice?

RQ3*:* In what ways can LLMs be designed to better meet the needs and preferences of diverse healthcare user groups?

RQ4*:* What future research directions should be pursued to ensure that LLMs are developed and implemented in a way that prioritizes human-centered design?

The findings from our scoping review highlight three key themes: supporting healthcare providers, enhancing patient interaction and empowering communities through education and training. Through a comprehensive analysis of these areas, it was evident that LLMs are significantly reshaping healthcare interactions between human-human and human-technology. We found that these technologies are influencing patient outcomes and raising serious ethical considerations. By improving diagnostic accuracy (e.g. [36]), providing timely and empathic patient interactions (e.g. [40]), and enhancing educational tools for medical professionals (e.g.[35]), LLMs are positioned to play a crucial role in the future of healthcare. However, as these technologies advance, it is important to address several practical implications, ensuring that they complement human expertise while safeguarding patient privacy and data security. In addition to the themes, the findings reveal several opportunities for future studies within the HCI community. Enhancing the personalization of LLM interactions, making patient care more compassionate and effective by focusing empathetic design, conducting longitudinal studies and adopting cross-disciplinary approaches, developing ethical frameworks, investigating the experiences of marginalized communities and improving the explainability of LLMs using transparent and comprehensible visualizations are some of them. We stress that it is critical to ensure the future work focuses not only on providing accurate results but also delivering culturally sensitive, and empathetic prompts.

This review contributes to HCI research by providing a clear roadmap for future research. It emphasizes the need for personalization, empathy, ethical considerations, inclusivity, and transparency in the development of LLMs for healthcare. As this is one of the first extensive scoping reviews, predominantly focused on the HCI literature, we believe that it lays the groundwork as a foundational reference to guiding future studies within the HCI community. Addressing gaps identified in this paper could help build trust and transparency, empowering users, and ensuring they understand and trust the decision-making process, which is essential for sustainable, effective and ethical healthcare delivery.

2 METHODOLOGY

The purpose of this study is to explore how LLMs are being used and discussed in the HCI literature, especially on the healthcare discourse. We have used the scoping review methodology [4] with PRISMA model [64, 87] as the methodology for this study. A scoping review is considered a thorough way to summarize existing research on a topic without evaluating the quality of each study. Particularly, it helps identify knowledge gaps and clarify concepts across a range of literature, making it a good starting point for understanding a field [43]. Unlike traditional reviews that critically evaluate evidence, scoping reviews focus on mapping out what has been studied [43]. A thematic analysis [5] of the selected studies was conducted to identify emerging themes. This process involved reading through the full texts to identify recurring themes related to the human-centered use of LLMs in healthcare

2.1 Eligibility Criteria

Studies were included if they satisfied the following eligibility criteria:

EC 1*: Use Large Language Models or Generative AI as a technology.*

EC 2*: Focuses on the topic of health or healthcare:*

- *includes physical and mental health topics,*
- *excludes literature that does not have health or healthcare as an essential part.*

***EC 3**: Was published or to be published after 2022 in HCI related proceedings or journals (Note that the papers focusing entirely on the computational improvements of AI algorithms such as accuracy improvements, were removed manually as the focus of this paper is to investigate how humans can better interact with LLMs in healthcare)*
***EC4**: Written in English language*.

2.2 Source & Terms

The primary source was the ACM Digital Library (ACM DL) as it is where most of the HCI related work is published [54, 88]. We supplemented the result by searching through Google Scholar to find the latest work that is in pre-prints [89] within the ACM digital library. This ensured a comprehensive examination of relevant studies. The exact query used was Abstract:(health* AND (ChatGPT OR "Generative AI" OR LLM* OR "Large Language Model")) OR Title:(health* AND (ChatGPT OR "Generative AI" OR LLM* OR "Large Language Model")). The asterisk (*) wildcard was used for the "health" term to ensure zero or more unknowns' characters were accepted which included terms such as healthcare in our search. The phrase E-Publication Date: (01/01/2022 TO *) was used to ensure we get the papers that were published from January 2022.

The final refined search phrase was;

Abstract:(health* AND (ChatGPT OR "Generative AI" OR LLM* OR "Large Language Model")) OR Title:(health* AND (ChatGPT OR "Generative AI" OR LLM* OR "Large Language Model")) AND E-Publication Date: (01/01/2022 TO *)

2.3 Screening

A CSV was computationally generated using the BIB file downloaded from the ACM digital library. And then the files found from the Google Scholar search were added to the CSV. After this, authors 1 and 2 screened the CSV individually against the eligibility criteria. Figure 1 below, illustrates the complete screening process.

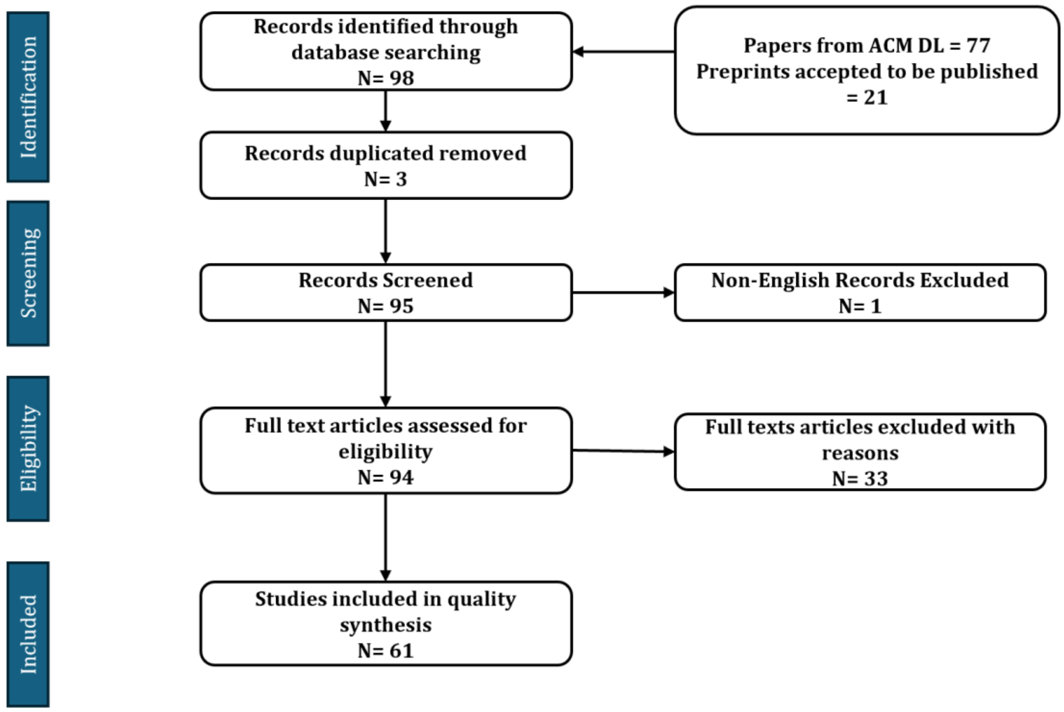

Figure 1 : The PRISMA flow diagram for the scoping review screening process [64]

After executing the search on the ACM digital library, we received 77 papers as of 21st of June 2024. We added an additional 21 papers collected from Google Scholar which were papers accepted to be published in the ACM digital library and were available as preprints at the time. The authors carefully went through all records and found three duplicated papers which were removed. Another one paper was excluded as it was a non-English paper. Only the title and abstract were written in English. The eligibility was assessed on the resulted paper set (n=94). There were 33 papers excluded as both author 1 and 2 agreed that the focus was only to improve the computational efficiency of the algorithm. These 33 papers had no evidence of any focus in enhancing the human element as was the focus of this paper. After going through the full text to double check the focus of the paper and ensure it is within the eligibility criteria, a total of 61 papers were included in our scoping review.

3 FINDINGS

The use of LLMs in healthcare is a rapidly evolving field with significant implications for patient care, support, and overall user experience [72]. This section explores the diverse themes emerging from recent literature from a HCI perspective, where the primary focus is on the users rather than the development or evaluation of AI models. The following themes have been identified: Supporting Healthcare Providers, Enhancing Patient Interaction, Empowering Communities through Education and Training. Through a comprehensive analysis of these themes, this section aims to provide a nuanced understanding of how LLMs are reshaping healthcare interactions, influencing patient outcomes, and raising critical ethical considerations.

3.1 Supporting Healthcare Providers (RQ2)

There is a branch of research which has explored the effectiveness of LLMs and LLM powered tools in providing diagnostic reports to assist healthcare professionals in making better decisions (e.g. [18, 20, 31]). Koopman and Zuccon [31] reports that ChatGPT, a well-known LLM, demonstrated an accuracy rate of 80% when answering health-related questions. They suggest that LLMs can assist healthcare professionals by offering reliable diagnostic support, particularly in situations where rapid decision-making is critical. Sayin et al. [50], complements this by reporting that LLMs could enhance medical decision-making processes by providing accurate and reliable suggestions. They claim that healthcare providers believe LLMs can help them to reduce the number of errors in diagnostics. Rajashekar et al. [46] have explored the integration of LLMs in clinical decision support systems (CDSS) for managing upper gastrointestinal bleeding (UGIB). The authors developed GutGPT, an AI-CDSS designed for risk prediction and guideline adherence in UGIB management. The study identified limitations such as the simulation lacking a real clinical environment for validation. Practical implications suggest that LLM-augmented AI-CDSS can enhance usability and accuracy in clinical settings by emphasizing their potential to support clinical decision-making. As such, researchers agree [31, 46, 50] that LLMs can help medical practitioners to enhance patient safety as accurate diagnostics are essential for effective treatment plans, and minimizing errors can lead to better health outcomes and potentially save lives. Additionally, researchers have explored the effectiveness of LLM powered tools in facing formal examinations. ChatGPT's performance on the United States Medical Licensing Examination (USMLE) which was evaluated and found that ChatGPT nearly reached the passing threshold without any specialized medical training, indicating its potential utility in clinical reasoning and medical education [35]. They suggest that LLMs can be further developed to assist in medical education and support in many accepts, LLMs can break language barriers by providing medical education and support in multiple languages. However, researchers state that it is crucial to note that these models are not infallible and require human oversight to ensure the accuracy and appropriateness of their recommendations.

LLMs have been utilized to predict patient outcomes based on diverse types of medical data, including electronic health records (EHRs), clinical notes, lab results, imaging reports, patient histories and clinical guidelines [34, 45, 78, 80]. These researchers have used electronic health records to train the LLMs and forecast disease progression and treatment responses. They found that LLMs could consistently monitor patient health data, provide timely alerts, and suggest appropriate interventions, which resulted in improving long term health outcomes. Kumar, Gattani and Singh [34] state that this would help healthcare providers to tailor treatments based on the specific needs of each patient. The real time recommendation generation of LLMs have also been explored to investigate how LLMs could support surgeons during procedures [50]. They report that LLMs could suggest optimal surgical techniques, assisting surgeons in making critical decisions by analyzing real-time data from various sources, including patient monitors, imaging equipment, and surgical instruments.

HealAI [18] is a tool designed specifically for medical documentation. The authors used prompt engineering and training optimized LLMs to create a cost-effective and customizable solution for healthcare documentation reducing the administrative burden on healthcare providers. [12, 61], explored the integration of LLMs and IoT technologies in healthcare to improve clinical decision-making and reduce medical errors. They highlight the need for comprehensive evaluation methods to ensure that such integrations enhance clinical decision-making and patient outcomes. It is found that LLMs have the potential to automate mundane routine tasks carried out by healthcare providers allowing them to focus on more complex decision making processes [1, 12, 18]. Jo et al. [27] explores the deployment of LLMs in public health interventions through a case study of the CareCall system. They report that the intervention helped reduce loneliness and offloaded workload from teleoperators. However, they also state the users expect the systems to handle emergency situations and provide direct social services which often exceed the capability to of it which could be dangerous without human intervention. Yildirim [77] reports that clinicians prefer AI for routine tasks over complex decision-making, indicating a potential role for LLMs in supporting radiology. However, [61] highlight the importance of addressing privacy and ethical concerns when integrating data from personal IoT devices such as smart watches when conducting further research.

The ethical considerations surrounding the use of LLMs in healthcare are also important to be addressed. This includes understanding how these models can be seamlessly incorporated into clinical workflows and ensuring that they complement rather than replace human expertise. Surani and Das [60] provide an in-depth analysis of privacy and security concerns associated with healthcare chatbots. They highlight several concerns from HCI perspective regarding the integration of LLMs in healthcare. The primary issue is managing user consent. While it is widely acknowledged that users must be fully informed about how their data will be used and that obtaining their explicit consent is essential, this is often not effectively implemented. Another issue is that users often face difficulties in exercising their right to have their data deleted. The processes for requesting data deletion are frequently unclear or cumbersome. Trust levels among users vary significantly, with older users generally expressing more concern about data privacy compared to younger users [19]. Building trust is crucial for the widespread adoption of healthcare chatbots [59, 60].

The complementary findings of Zhengliang Liu et al. [39] , state that LLMs can assist doctors in making more informed decisions while retaining the essential human judgment and oversight in the diagnostic process. They also address ethical concerns related to the use of AI in medical diagnostics, such as patient privacy, data security, and the potential for bias in AI-generated diagnoses and claim that building trust with both medical professionals and patients is crucial for the successful integration of AI in healthcare. Several studies [2, 25, 36, 82] mirror this sentiment by reporting it's crucial to collaborate with regulatory bodies to develop clear guidelines focusing on ensuring patient safety, data privacy, and the ethical use of AI in healthcare. Research by Sayin et al. [50] emphasize the importance of maintaining transparency in AI decision-making processes. Ensuring that healthcare professionals understand how LLMs arrive at their recommendations is crucial for building trust and facilitating effective collaboration between humans and AI.

3.2 Enhancing Patient Interactions (RQ 1)

LLM-powered tools have been found to improve patient engagement by providing timely responses to health-related queries and offering emotional support to users [68]. In this section we analyze previous research that explored how LLM powered tools interact with patients.

There is a set of research available that provides insights into the application of LLMs in various health categories of patients particularly diabetes prevention and management [11], skin issues [37], mental health [44, 56, 73, 75], elderly health [3] and child health [51, 62]. Koulouri et al. [32] have explored the acceptability and effectiveness of using chatbots as mental health support tools for young adults. They conduct user interviews and surveys with a diverse group of young adults to gather insights on their experiences and perceptions of mental health chatbots and analyze the interaction logs of chatbot usage. They report that LLM powered chatbots proved to be effective in offering immediate support, particularly during crisis situations, by providing coping strategies and resources. Most users found chatbots to be an acceptable form of mental health support, appreciating their availability and ease of access. Preferred features included the anonymity they offered, the quick responses, and the ability to express thoughts without fear of judgment. However, some of the participants have raised some concerns about the limitations of chatbots in understanding complex emotional states and providing personalized responses [45]. A complementing study was conducted by You et al. [78] where they investigate the effects of chatbot-based symptom checkers' conversational design on user behavior and emotions, and to identify ways to enhance user interactions with these apps. They have used interviews and experimental studies to explore potential conversational design solutions based on the interview findings. They report that providing emotional support through conversational support chatbots positively impacts user perceptions and experiences. Clear and detailed explanations of medical information enhance user trust and aid in decision-making. They state that these tools could significantly impact user satisfaction through emotional support and explanations and help patients manage their health better and reduce anxiety about symptoms or conditions. However, most current discussions on trust in LLM powered tools often neglect the user's perspective, focusing instead on the principles of trustworthy AI without considering how users form trust judgments [38]. Koulouri et al. [32] state that the effects of emotional support and detailed explanations on user perceptions can vary depending on the context of the interaction. Telemedicine has gained significant traction, especially in the wake of the COVID-19 pandemic. These virtual assistants can handle a wide range of tasks, from answering general health queries to conducting preliminary assessments of symptoms. According to You et al. [78], these AI-driven virtual assistants can respond to patients, providing guidance on whether they should seek in-person medical care or manage their condition at home. This capability is particularly valuable in rural or underserved areas where access to healthcare professionals may be limited. By offering round-the-clock support, virtual health assistants ensure that patients receive timely information and care, improving overall healthcare accessibility. Yang et al. [76], also examines the impact of LLM-based Telehealth Voice Assistant for Older Adults on patient-provider communication, leading to better healthcare outcomes and increased patient satisfaction. The authors conducted interview studies with older adults and healthcare providers to assess the effectiveness of the voice assistant. Wang et al. [69], highlight the importance of mobile health (mHealth) applications in patient-provider communication, specifically for HIV care by improving patient engagement and communication quality in healthcare. They also highlight that LLMs can significantly enhance the emotional and affective quality of mHealth communication, improving patient engagement and satisfaction by providing more personalized and empathetic interactions. Another study [37] explores the design and implementation of AI-based conversational agents for healthcare applications. They employed a hybrid Wizard-of-Oz study design [16], combining human operators and AI responses to simulate a conversational agent. They report that participants found these tools helpful in understanding their medical situation and alleviating their concerns. Most participants' experiences were positive in interacting with the tool. Researchers agree that improvements are needed to address emotional and empathetic interactions in modern LLM powered tools used in healthcare [27, 52].

Alessa and Hend Al-Khalifa [3] investigate the role of LLM powered chatbots in enhancing elderly health and providing social support by providing an overview of the system's inspiration, functionality, design and focusing on creating a user-friendly interface. They highlight the importance of designing a conversational companion tailored to the needs of elderly users, emphasizing the importance of user-specific data and continuous refinement as ChatGPT effectively enhances social support for elderly individuals. Soun and Nair [56], analyzed GPT 3.5 based mental health applications. The study exhibits notable biases, particularly towards young female patients. These ethical concerns prompt the need for continuous evaluation and improvement of AI models to ensure fair and effective mental health support [73].

Tang et al. [62] have evaluated an emotional learning tool designed for children with High-Functioning Autism (HFA). The findings indicate that HFA children showed significant improvements in emotional recognition and expression after using the tool for ten days. Children completed an average of 52.5 conversations, demonstrating high engagement and usability of the tool. The study also highlighted the potential of GenAI in supporting personalized emotional learning. Stapleton et al. [57], explore and address the challenges faced by online volunteers in supporting individuals with suicidal ideation. Researchers conducted interviews using role-playing conversations to simulate scenarios with varying severities of suicidality. The study employed a "speed dating" approach to present several design concepts to participants, focusing on emotional preparation and support, real-time guidance, training, and suicide detection, aiming to gather feedback on potential solutions. While the researchers saw potential in AI-based technologies for training and guidance, they emphasized the need for human discretion and contextual understanding. The volunteers expressed a strong desire for more emotional preparation and support mechanisms, highlighting the importance of peer support [79] and resources to manage the emotional stress of their work. It was also found when interacting with LLM powered conversational agents, that children perceived them as a peer and felt comfortable sharing both positive and negative experiences [52]. Authors report that some children were disappointed by the systems' limitations in sharing its own experiences or emotions and emphasize the need for more sophisticated interaction capabilities.

3.3 Empowering communities through Education and Training (RQ 3)

Studies are conducted to explore the experiences of integrating LLMs into healthcare systems in various marginalized communities. Najeeb et al. [1] explore the role of LLMs and conversational agents (CAs) in healthcare peer support programs by providing consistent and reliable information, facilitating communication among participants. The authors employed an ethnographic approach to understand the facilitators' needs and the potential of LLMs to support peer support programs. However, as this study was done as the solitary case study in Kenya, it may not be universally applicable. Yadav et. [74] explored the potential of using chatbots to educate women about breastfeeding in urban areas of Delhi, India. The study involved a Wizard-of-Oz experiment with twenty-two participants, including breastfeeding mothers and community health workers. The findings showed that the majority of questions from users could be addressed by the chatbot, indicating its potential as a valuable tool for providing reliable breastfeeding information. The study also highlights the significant influence of female relatives on breastfeeding practices and the need for chatbot designers to focus on positive reinforcement and contextual sensitivity. Steenstra et al. [58] suggest improving accessibility of digital forms for older adults and marginalized groups when entering health data by highlighting the importance of developing digital solutions that cater the needs of overlooked demographics, particularly through multimodal interfaces to enhance the usability for older adults. Najarro et al. [6] introduces WMGPT, a system designed to provide mental health counseling using a customized GPT model. It was aimed to counter the limitations of traditional in-person counseling such as social stigma.

Self-reflection is an aspect of mental health that helps person understand about themselves, both young and old [35, 52]. It encourages a person to better understand their own experiences and perspectives. When this is applied within a community, it can lead to collective understanding where each community member undiscussed each other's struggles as well as triumphs. Research has been conducted [30] to explore how LLM driven journaling apps can be used to help psychiatric patients in better understanding their thoughts and daily contexts. Kim et al. [30] report that MindfulDiary, an LLM driven diary application helped patients to move deeper into their thoughts. Patients report that they felt a sense of empathy from the application which helped them maintain consistent journaling habits. Another study has [44] extended this work by integrating LMS powered journaling tool with the work of college students targeting their well-being. They plan to investigate how such a system can enhance mental well-being through context-aware self-reflection.

Another aspect of empowerment where LLMs attempted to tap into is mental resilience. Communities often encounter various crisis situations (natural disasters, economic downturns, or pandemics) and mentally resilient individuals are better equipped to handle stress and recover from adverse situations [17, 65]. Hu et al. [22] investigates the potential of using LLMs to cultivate mental resilience in children. Authors employed qualitative interviews with children and their parents where they used a mobile app powered with an LLM designed to enhance psychological resilience. The overall feedback was positive on usability and effectiveness. However, the authors highlight the need to balance educational and engaging content. While engagement kept the children using the system, educational content ensured that the interactions are meaningful and beneficial for their mental resilience. Also, since children have shorter attention spans [63, 90], it is also crucial to optimize the interaction length and speed. These will help ensure sustained interest, effective learning, and a positive user experience.

Researchers have investigated the use of virtual agent-based games to enhance health education among adolescents [58]. The authors developed a fantasy game using LLMs to enhance health knowledge and vaccine acceptance among adolescents. The study shows that games can improve learning and engagement in health education, particularly among adolescents. However, they highlight further diverse studies are needed to confirm these findings. Wu et al. [71], explores the use of LLMs at reducing problematic smartphone use. They used a LLM powered tool named MindShift which considers users' in-the-moment behaviors, physical contexts, mental states, goals, and habits to generate personalized persuasive content. They developed four strategies: understanding, comforting, evoking, and scaffolding habits, each focused on different mental states (boredom, stress, inertia). The study revealed that the MindShift intervention significantly reduced smartphone usage duration by 7.4-9.8%, indicating less time spent on devices. Participants also reported notable improvements in self-control. User feedback further highlighted the effectiveness and acceptability of Mindshift's personalized and context-aware interventions, which were preferred over static, repetitive reminders. Ding et al. [14], have looked into how language patterns in social media discussions on public health practices can be analyzed using LLMs which can influence national health outcomes. The study shows that people's engagement with health-related discussions on social media can influence and reflect real-world health outcomes. This highlights the importance of understanding and monitoring online discourse to guide public health strategies. The authors report that social media play a critical role in influencing health behaviors and outcomes, demonstrating the potential of LLMs in analyzing and leveraging online discussions for public health intervention. Elise et al. [29] complements this finding by experimenting how LLMs can create persuasive public health messages to invoke positive emotions towards vaccinations. It was found that LLM generated messages were perceived as more effective, stronger arguments, and invoked more positive attitudes than those created by the CDC (Centers for Disease Control and Prevention) USA. However, it was also reported that despite the quality of LLM generated content, individuals trusted messages from the CDC more than those labeled as AI-generated. This study and many others [50, 60] highlight a critical aspect of trust in communication. While LLMs can produce compelling messages, users still place higher trust in content from recognized human institutions.

4 FUTURE DIRECTIONS FOR THE INTEGRATION OF LLMS IN HEALTHCARE: HUMAN-CENTERED APPROACHES (RQ 4)

The use of LLMs in healthcare offers a set of opportunities to support patient care, streamline healthcare services, and assist medical professionals. We believe these benefits can be maximized by prioritizing a human-centered approach, focusing on usability, ethical considerations, and collaborative interdisciplinary efforts. This section presents such future research directions focusing on not only to deliver advanced technological solutions but also meet the diverse needs and concerns of patients and healthcare providers, ultimately enhancing patient care and improving healthcare outcomes.

4.1 Personalized Interactions

Personalized and context-aware interactions have long been of interest within CHI communities around the world. They can significantly enhance user experience and efficiency by tailoring content, functionality, and design to individual preferences [11], For instance, in healthcare, personalized versions of ChatGPT have shown varied accuracy in health information provision based on the specificity of user prompts, underscoring the importance of tailored interactions [31]. In the field of medical education, AI systems such as ChatGPT have demonstrated near-passing scores on the USMLE without specialized training, highlighting their potential to offer personalized learning experiences [35].

However, research on improving personalization in LLMs is still in its initial stages. As such we believe that future researchers should prioritize providing advice that is not only relevant and precise but also aligns closely with individual needs, significantly boosting user satisfaction and trust. You et al. [78] discuss the significant benefits of such personalized interactions, noting how they can deepen user trust and enhance the overall experience.

Incorporating personal health data into LLM interactions is crucial for offering tailored advice and recommendations [12]. The emergence of wearable and IoT devices present a great opportunity with this regard [49, 77]. As such, future research should focus on developing methods to effectively integrate data from wearable devices such heart rate monitors, glucose sensors, and fitness trackers into LLM interactions. This integration enables LLMs to provide personalized health advice tailored to the individual's real-time physiological data [11]. also highlight the necessity for personalized health management plans that include custom health goals and behavior interventions tailored to user habits and preferences.

The benefits of such personalized interfaces are multifaceted. On the one hand, the patients would get more relevant and precise medical advice, leading to greater engagement and adherence to treatment plans. On the other hand, health care professionals are enabled to deliver more effective care with tailored information, making better-informed decisions that cater specifically to each patient's unique health conditions and preferences. Systems that adjust to individual patient needs which can reduce unnecessary treatments and streamline processes, potentially lowering costs and improving resource allocation.

By focusing on personalization and integrating advanced technologies such as IoT, HCI researchers can help ensure that AI applications in healthcare are not only efficient but also compassionate and patient-centered, leading to better health outcomes and higher satisfaction among all stakeholders involved.

4.2 Longitudinal and cross disciplinary interventions

The CHI community has long engaged in discussions about the balance between understanding the *long-term implications* of technology use versus focusing on *short-term explorations* of emerging tools and interfaces. As such another area that remains underexplored is the long-term impacts and implications of these LLM interventions. To fully understand and harness the potential of AI in healthcare, it is essential to conduct longitudinal studies and embrace cross-disciplinary approaches. These studies should allow researchers to track changes in patient outcomes, adherence to treatments, and overall healthcare experiences over time [31]. For instance, while initial studies may show high accuracy rates of AI in diagnostics, longitudinal studies can reveal how these tools influence patient behavior, long-term health outcomes, and the evolution of disease management practices. Moreover, the dynamic nature of AI technologies necessitates continuous evaluation. Longitudinal studies can help in understanding how updates and improvements to AI systems affect their performance and user interaction. By tracking the evolution of AI systems and their interaction with patients and healthcare providers, longitudinal studies help ensure these technologies remain relevant, effective, and aligned with clinical needs.

In addition to longitudinal studies, we believe cross-disciplinary approaches and disciplinary approaches are vital in addressing the diverse challenges associated with integrating LLMs in healthcare. Researchers agree that addressing the complex challenges of integrating LLMs in healthcare would benefit from insights from multiple disciplines [10, 83]. Integrating HCI expertise with AI technologies can significantly improve the usability and effectiveness of LLMs in healthcare. Future research initiatives could focus on developing LLMs that adhere to user-centered design principles, ensuring that the systems are tailored to meet the specific needs, preferences, and experiences of end-users by making users a central part of the design process. This complements Zhu et al. [81] where they advocate for iterative prototyping and usability testing with both healthcare providers and patients. This multidisciplinary collaboration is necessary to design user-centered tools, develop robust data privacy and security measures, and create ethical frameworks for responsible use of LLMs in healthcare [10, 25, 78]. Such a holistic understanding would greatly benefit end users. For example, healthcare providers would gain from interfaces that are intuitive and streamline complex decision-making processes. This will help reduce cognitive load and increase focus on patient care. Patients benefit from more engaging and understandable interactions with AI-driven health applications, which can enhance their overall experience and satisfaction with healthcare services. Additionally, the AI researcher would gain valuable insights from user feedback that can guide the optimization of existing LLM functionalities. This will ensure the technologies developed are not only advanced but also practical and user-friendly. As such, we believe that adopting a longitudinal and collaborative approach will not only drive innovation but also ensure that the technology developed is both effective and well-suited for sustainable use in real-world healthcare applications in the long term.

4.3 Respectful and responsible design

The healthcare sector is known for its intricate regulatory and ethical frameworks, which adds complexity to the development and deployment of LLMs. Additionally, one of the major stakeholders - patients are often vulnerable. This makes it even more crucial to incorporate ethical and legal insights at every stage of LLM development to ensure these technologies align with healthcare standards and prioritize patient welfare. Researchers agree that it is important to have comprehensive ethical and legal frameworks to govern the use of LLMs in healthcare [55, 66, 67]. These frameworks would ensure that AI systems are used responsibly, protecting patient rights and maintaining public trust. It’s also important not to simply build trustworthy applications, but also to effectively communicate this trustworthiness to the users [38] .

Considering these, we suggest that future research should aim to develop ethical frameworks that guide the responsible use of AI in this field, ensuring that considerations such as patient rights, data privacy, healthcare regulations and standards, and fairness are at the forefront. We suggest that researchers should focus not only on being innovative and effective but also on being responsible and respectful of the unique demands of the healthcare sector. It should be noted that respectful and responsible design [53] is more than adhering to rules or laws; it is about acknowledging human dignity and individual characteristics, treating people as ends in themselves rather than means to an end. As such, these frameworks need to be crafted in collaboration with a diverse group of stakeholders, including healthcare professionals, ethicists, and legal experts. Such collaboration ensures that the guidelines are well-rounded and practical, addressing the varied perspectives and concerns of those involved in healthcare delivery.

Another potential research direction would be to explore how to get patients and healthcare providers involved in the design process of LLMs [67]. This may include incorporation of conducting user studies, focus groups, and surveys to gather feedback and iteratively improve the models. This requires integrating comprehensive patient data securely and ethically, ensuring that the individuality of each patient is respected. Furthermore, we should explore how these systems can prioritize obtaining informed consent from patients before using their data or making decisions on their behalf. Exploring how to ensure patient autonomy by presenting recommendations as options, and not as final decisions would also be interesting. A comparative study where the recommendations are given for one set as options and the other as final decisions could generate valuable insights. LLMS can also be designed to exhibit empathetic behaviors, understanding the emotional and psychological needs of patients. For example, virtual assistants powered by LLMs can provide comfort and reassurance to patients undergoing stressful medical procedures.

Inspired by studies in Kenya [1] and rural India [74] , another interesting avenue of research would be to explore the experiences of users from marginalized communities in using LLMs to manage health related activities. Marginalized communities often face a specific set of challenges when accessing healthcare and technology. Infrastructure issues, low digital accuracy, language barriers, skepticism towards modern technology and western medical practices, and social stigmas are some of these challenges. While many of the interventions within the western society resulted in largely positive outcomes, it would be interesting to investigate whether the same study would produce comparable results with these marginalized communities. Localizing these LLMS to suit and address these specific challenges could be another future research.

4.4 Emotional Support and Empathetic Design

Emotional support [32, 78] and positive feedback [8] in LLM interactions are found to be positively correlated with better health outcomes. However, it was also found that managing conversations involving suicidal ideation presents significant challenges [57].

Accordingly, future research can focus on identifying factors that contribute to positive outcomes aiming to enhance the impact of LLM involvements in mental health interventions. These factors can then be incorporated into LLMs to provide more empathetic and supportive interactions, which are essential for mental health applications. As [8] found that positive feedback correlates with better mental health outcomes, future researchers can explore how these positive reinforcement can be applied in LLMs to improve user engagement.

Studies have also shown that LLMs have the potential to positively influence the emotional learning of younger users [28, 62] . As such it would be interesting to study how LLMs can be used to better support emotional development and how they can help young users better understand and manage their emotions, contributing to healthier psychological growth.

Reinforcing LLMs with emotional intelligence, which would enable more sympathetic replies, is another possible research direction. One interesting topic of study would be to investigate how these empathetic reactions affect patients' ability to manage stress and worry connected to their health. Additionally, exploring how healthcare providers leverage these emotionally intelligent systems to gain deeper insights into their patients' emotional and psychological states would help better design these tools for the future. Another possibility would be to explore how LLMs should handle such sensitive topics to provide immediate crisis support and ensure safe escalation to human professionals. The sustained interaction of these systems, while protecting vulnerable users and maintaining their interest and engagement over time, is also a key area for future research.

4.5 Improving Explainability for Users through effective visualizations

Aligned with the suggestions of [31, 83], who advocate for making AI's decision-making process more transparent and comprehensible, another critical future direction is to improve the explainability of these models for end users. Transparency is particularly important because people naturally want to understand what they are engaging with and why certain outcomes occur. Transparency would help to build trust between the technology and its users, enabling patients and healthcare providers to understand how LLM powered systems arrive at their conclusions and feel confident in using them. It will also enable the healthcare sector to ensure accountability, allowing healthcare providers to evaluate the appropriateness of the recommendations provided and take corrective action if necessary.

As such, future research could focus on ways to improve the explainability aspects of LLM results. Research can explore development of algorithms that offer clear, straightforward explanations of their outcomes and designing user interfaces that allow users to see and understand the logic behind provided recommendations. Participatory design studies and focus group sessions can help identify how laypersons perceive the outcomes of certain models. This will help guide the design of systems that communicate effectively with diverse user groups. Instead of relying on a one-size-fits-all approach to visualization techniques for explaining these algorithms, researchers can explore the most optimal visualization techniques tailored to each specific scenario. The effectiveness of visualizations such as flowcharts, heatmaps, and decision trees which have the potential to transform complex algorithms into understandable graphics, must be evaluated using techniques such as A/B testing [91]. These research findings can be used to measure effectiveness of such visualizations in improving the explainability. Collaboration with healthcare professionals can ensure the explanations provided by LLM powered systems are relevant and useful in a clinical context, while educational initiatives and regulatory frameworks can further support transparency. By enhancing explainability, these technologies become not only more transparent but also more trustworthy, leading to wider acceptance and more effective implementation in healthcare settings.

5 CONCLUSION

The rapid growth of LLMs and related fields are revolutionizing several fields of study including healthcare. Particularly in healthcare, this integration can be used to impact aspects such as patient care, healthcare delivery, and supporting medical professionals in decision making. However, while much of the previous research is carried out to explore the improvements of computational efficiencies and result generation accuracy, the work focusing on supporting the human experience when interacting with such LLM powered tools are lacking. As such, this integration of LLMs in the healthcare sector presents a new frontier of research for HCI practitioners. The aim of this paper was to synthesis the HCI work related to healthcare sector to provide a foundation for future research.

Accordingly, we have conducted a scoping review using the PRISMA model focusing mainly on the HCI research in incorporating LLMs in healthcare. Through a thematic analysis, this paper highlights the roles LLMs can play, from providing diagnostic support and personalized health advice to offering emotional and mental health support. Our analysis reveals three key themes, including supporting healthcare providers, enhancing patient interactions, and empowering communities through education and training. Additionally, the review identifies significant opportunities for future research from HCI perspective, emphasizing the importance of personalized interactions, longitudinal and cross-disciplinary studies, respectful and responsible design, emotional support, empathetic design, and improving the explainability of AI systems.

Through this paper, we make the case that by prioritizing a human-centered approach, future research can ensure that LLMs are developed and implemented to meet the needs and preferences of diverse patient groups and healthcare providers. We suggest that such an approach will not only help advance technological innovation but also promote ethical, effective, and sustainable use of technology in healthcare. We also promote the use of collaborative efforts among HCI researchers, AI developers, healthcare professionals, and regulatory bodies to address the complex challenges associated with integrating LLMs in healthcare. We highlight how the successful integration of LLM powered tools requires careful consideration of human factors, ethical standards, and continuous evaluation to ensure they complement human expertise and safeguard patient privacy and data security. This comprehensive scoping review contributes to the HCI research by serving as a foundational reference for future research in the HCI community, guiding the development of LLMs that are not only advanced and efficient but also compassionate, ethical, and centered around the users they aim to serve.

REFERENCES

[1] Najeeb Gambo Abdulhamid, Millicent Ochieng, Kalika Bali, Elizabeth Ankrah, Naveena Karusala, Keshet Ronen, and Jacki O'Neill. 2023. Can Large Language Models Support Medical Facilitation Work? A Speculative Analysis. In *ACM International Conference Proceeding Series*, November 2023. Association for Computing Machinery, 64–70. https://doi.org/10.1145/3628096.3628752

[2] Khadija Alam, Akhil Kumar, and F. N. U. Samiullah. 2024. Prospectives and drawbacks of ChatGPT in healthcare and clinical medicine. *AI and Ethics* (February 2024). https://doi.org/10.1007/s43681-024-00434-5

[3] Abeer Alessa and Hend Al-Khalifa. 2023. Towards Designing a ChatGPT Conversational Companion for Elderly People. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, July 05, 2023. ACM, Corfu Greece, 667–674. https://doi.org/10.1145/3594806.3596572

[4] Hilary Arksey and Lisa O'Malley. 2005. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 8, 1 (February 2005), 19–32. https://doi.org/10.1080/1364557032000119616

[5] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.*, Harris Cooper, Paul M. Camic, Debra L. Long, A. T. Panter, David Rindskopf and Kenneth J. Sher (eds.). American Psychological Association, Washington, 57–71. https://doi.org/10.1037/13620-004

[6] Lismer Andres Caceres Najarro, Yonggeon Lee, Kobiljon E. Toshnazarov, Yoonhyung Jang, Hyungsook Kim, and Youngtae Noh. 2023. WMGPT: Towards 24/7 Online Prime Counseling with ChatGPT. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, October 08, 2023. ACM, Cancun, Quintana Roo Mexico, 142–145. https://doi.org/10.1145/3594739.3610708

[7] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology* (January 2024). https://doi.org/10.1145/3641289

[8] Prerna Chikersal, Danielle Belgrave, Gavin Doherty, Angel Enrique, Jorge E. Palacios, Derek Richards, and Anja Thieme. 2020. Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention. In *Conference on Human Factors in Computing Systems - Proceedings*, April 2020. Association for Computing Machinery. https://doi.org/10.1145/3313831.3376341

[9] Kenneth Mark Colby. 1981. Modeling a paranoid mind. *Behav Brain Sci* 4, 4 (December 1981), 515–534. https://doi.org/10.1017/S0140525X00000030

[10] Sebastian Cross, Ahmed Mourad, Guido Zuccon, and Bevan Koopman. 2021. Search engines vs. symptom checkers: A comparison of their effectiveness for online health advice. In *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, April 2021. Association for Computing Machinery, Inc, 206–216. https://doi.org/10.1145/3442381.3450140

[11] Dung Dao, Jun Yi Claire Teo, Wenru Wang, and Hoang D. Nguyen. 2024. LLM-Powered Multimodal AI Conversations for Diabetes Prevention. In *Proceedings of the 1st ACM Workshop on AI-Powered Q&A Systems for Multimedia*, June 10, 2024. ACM, Phuket Thailand, 1–6. https://doi.org/10.1145/3643479.3662049

[12] Gabriele De Vito. 2024. Assessing healthcare software built using IoT and LLM technologies. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*, June 18, 2024. ACM, Salerno Italy, 476–481. https://doi.org/10.1145/3661167.3661202

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Retrieved September 27, 2024 from http://arxiv.org/abs/1810.04805

[14] Xiaohan Ding, Buse Carik, Uma Sushmitha Gunturi, Valerie Reyna, and Eugenia Ha Rim Rho. 2024. Leveraging Prompt-Based Large Language Models: Predicting Pandemic Health Decisions and Outcomes Through Social Media Language. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–20. https://doi.org/10.1145/3613904.3642117

[15] Yutao Dou, Yuwei Huang, Xiongjun Zhao, Haitao Zou, Jiandong Shang, Ying Lu, Xiaolin Yang, Jian Xiao, and Shaoliang Peng. 2024. ShennongMGS: An LLM-based Chinese Medication Guidance System. *ACM Trans. Manage. Inf. Syst.* (April 2024), 3658451. https://doi.org/10.1145/3658451

[16] World Leaders in Research-Based User Experience. The Wizard of Oz Method in UX. *Nielsen Norman Group*. Retrieved June 27, 2024 from https://www.nngroup.com/articles/wizard-of-oz/

[17] David Fletcher and Mustafa Sarkar. 2013. Psychological Resilience. *European Psychologist* 18, 1 (January 2013), 12–23. https://doi.org/10.1027/1016-9040/a000124

[18] Sagar Goyal, Eti Rastogi, Sree Prasanna Rajagopal, Dong Yuan, Fen Zhao, Jai Chintagunta, Gautam Naik, and Jeff Ward. 2024. HealAI: A Healthcare LLM for Effective Medical Documentation. In *WSDM 2024 - Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, March 2024. Association for Computing Machinery, Inc, 1167–1168. https://doi.org/10.1145/3616855.3635739

[19] David Goyeneche, Stephen Singaraju, and Luis Arango. 2023. Linked by age: a study on social media privacy concerns among younger and older adults. *Industrial Management & Data Systems* 124, 2 (January 2023), 640–665. https://doi.org/10.1108/IMDS-07-2023-0462

[20] Jiyeon Han, Jimin Park, Jinyoung Huh, Uran Oh, Jaeyoung Do, and Daehee Kim. 2024. AscleAI: A LLM-based Clinical Note Management System for Enhancing Clinician Productivity. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–7. https://doi.org/10.1145/3613905.3650784

[21] Richard Harper and Dave Randall. 2024. Machine Learning and the Work of the User. *Comput Supported Coop Work* 33, 2 (June 2024), 103–136. https://doi.org/10.1007/s10606-023-09483-6

[22] Zihui Hu, Hanchao Hou, and Shiguang Ni. 2024. Grow with Your AI Buddy: Designing an LLMs-based Conversational Agent for the Measurement and Cultivation of Children?s Mental Resilience. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, June 17, 2024. ACM, Delft Netherlands, 811–817. https://doi.org/10.1145/3628516.3659399

[23] Yining Hua, Fenglin Liu, Kailai Yang, Zehan Li, Yi-han Sheu, Peilin Zhou, Lauren V. Moran, Sophia Ananiadou, and Andrew Beam. 2024. Large Language Models in Mental Health Care: a Scoping Review. Retrieved June 26, 2024 from http://arxiv.org/abs/2401.02984

[24] Nchebe-Jah Iloanusi and Soon Ae Chun. 2024. AI Impact on Health Equity for Marginalized, Racial, and Ethnic Minorities. In *Proceedings of the 25th Annual International Conference on Digital Government Research*, June 11, 2024. ACM, Taipei Taiwan, 841–848. https://doi.org/10.1145/3657054.3657152

[25] Raquel Iniesta. 2023. The human role to guarantee an ethical AI in healthcare: a five-facts approach. *AI and Ethics* (October 2023). https://doi.org/10.1007/s43681-023-00353-x

[26] Yiqiao Jin, Mohit Chandra, Gaurav Verma, Yibo Hu, Munmun De Choudhury, and Srijan Kumar. 2024. Better to Ask in English: Cross-Lingual Evaluation of Large Language Models for Healthcare Queries. In *Proceedings of the ACM on Web Conference 2024*, May 13, 2024. ACM, Singapore Singapore, 2627–2638. https://doi.org/10.1145/3589334.3645643

[27] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung, and Young Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Conference on Human Factors in Computing Systems - Proceedings*, April 2023. Association for Computing Machinery. https://doi.org/10.1145/3544548.3581503

[28] Eunkyung Jo, Young-Ho Kim, Yuin Jeong, Sohyun Park, and Daniel A. Epstein. Incorporating Multi-Stakeholder Perspectives in Evaluating and Auditing of Health Chatbots Driven by Large Language Models. Retrieved from https://guide.ncloud-docs.com/

[29] Elise Karinshak, Sunny Xun Liu, Joon Sung Park, and Jeffrey T. Hancock. 2023. Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023). https://doi.org/10.1145/3579592

[30] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-Woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. 2024. MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling. May 2024. Association for Computing Machinery (ACM), 1–20. https://doi.org/10.1145/3613904.3642937

[31] Bevan Koopman and Guido Zuccon. Dr ChatGPT tell me what I want to hear: How different prompts impact health answer correctness. 15012–15022. Retrieved from https://github.com/

[32] Theodora Koulouri, Robert D. MacRedie, and David Olakitan. 2022. Chatbots to Support Young Adults' Mental Health: An Exploratory Study of Acceptability. *ACM Transactions on Interactive Intelligent Systems* 12, 2 (June 2022). https://doi.org/10.1145/3485874

[33] Harsh Kumar, Yiyi Wang, Jiakai Shi, Ilya Musabirov, Norman A. S. Farb, and Joseph Jay Williams. 2023. Exploring the Use of Large Language Models for Improving the Awareness of Mindfulness. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, April 19, 2023. ACM, Hamburg Germany, 1–7. https://doi.org/10.1145/3544549.3585614

[34] Rohit Kumar, Dr Ram Krishna Gattani, and Kavita Singh. 2024. Enhancing Medical History Collection using LLMs. In *Proceedings of the 2024 Australasian Computer Science Week*, January 29, 2024. ACM, Sydney NSW Australia, 140–143. https://doi.org/10.1145/3641142.3641174

[35] Tiffany H. Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, and Victor Tseng. 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health* 2, 2 (February 2023), e0000198. https://doi.org/10.1371/journal.pdig.0000198

[36] Kenya Kusunose, Shuichiro Kashima, and Masataka Sata. 2023. Evaluation of the Accuracy of ChatGPT in Answering Clinical Questions on the Japanese Society of Hypertension Guidelines. *Circulation Journal* 87, 7 (July 2023), 1030–1033. https://doi.org/10.1253/circj.CJ-23-0308

[37] Brenna Li, Amy Wang, Patricia Strachan, Julie Anne Séguin, Sami Lachgar, Karyn C Schroeder, Mathias S Fleck, Renee Wong, Alan Karthikesalingam, Vivek Natarajan, Yossi Matias, Greg S Corrado, Dale Webster, Yun Liu, Naama Hammel, Rory Sayres, Christopher Semturs, and Mike Schaekermann. 2024. Conversational AI in health: Design considerations from a Wizard-of-Oz dermatology case study with users, clinicians and a medical LLM. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–10. https://doi.org/10.1145/3613905.3651891

[38] Q. Vera Liao and S. Shyam Sundar. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. In *ACM International Conference Proceeding Series*, June 2022. Association for Computing Machinery, 1257–1268. https://doi.org/10.1145/3531146.3533182

[39] Zhengliang Liu, Aoxiao Zhong, Yiwei Li, Longtao Yang, Chao Ju, Zihao Wu, Chong Ma, Peng Shu, Cheng Chen, Sekeun Kim, Haixing Dai, Lin Zhao, Lichao Sun, Dajiang Zhu, Jun Liu, Wei Liu, Dinggang Shen, Xiang Li, Quanzheng Li, and Tianming Liu. 2024. Radiology-GPT: A Large Language Model for Radiology. Retrieved June 24, 2024 from http://arxiv.org/abs/2306.08666

[40] Man Luo, Christopher J. Warren, Lu Cheng, Haidar M. Abdul-Muhsin, and Imon Banerjee. 2024. Assessing Empathy in Large Language Models with Real-World Physician-Patient Interactions. (May 2024). Retrieved from http://arxiv.org/abs/2405.16402

[41] Xiangbin Meng, Xiangyu Yan, Kuo Zhang, Da Liu, Xiaojuan Cui, Yaodong Yang, Muhan Zhang, Chunxia Cao, Jingjia Wang, Xuliang Wang, Jun Gao, Yuan Geng Shuo Wang, Jia ming Ji, Zifeng Qiu, Muzi Li, Cheng Qian, Tianze Guo, Shuangquan Ma, Zeying Wang, Zexuan Guo, Youlan Lei, Chunli Shao, Wenyao Wang, Haojun Fan, and Yi Da Tang. 2024. The application of large language models in medicine: A scoping review. *iScience* 27, 5 (May 2024). https://doi.org/10.1016/j.isci.2024.109713

[42] Sara Montagna, Stefano Ferretti, Lorenz Cuno Klopfenstein, Antonio Florio, and Martino Francesco Pengo. 2023. Data Decentralisation of LLM-Based Chatbot Systems in Chronic Disease Self-Management. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, September 06, 2023. ACM, Lisbon Portugal, 205–212. https://doi.org/10.1145/3582515.3609536

[43] Zachary Munn, Micah D. J. Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 18, 1 (December 2018), 143. https://doi.org/10.1186/s12874-018-0611-x

[44] Subigya Nepal, Arvind Pillai, William Campbell, Talie Massachi, Eunsol Soul Choi, Xuhai Xu, Joanna Kuc, Jeremy F Huckins, Jason Holden, Colin Depp, Nicholas Jacobson, Mary P Czerwinski, Eric Granholm, and Andrew Campbell. 2024. Contextual AI Journaling: Integrating LLM and Time Series Behavioral Sensing Technology to Promote Self-Reflection and Well-being using the MindScape App. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–8. https://doi.org/10.1145/3613905.3650767

[45] Joel Oduro-Afriyie and Hasan M Jamil. 2023. Enabling the Informed Patient Paradigm with Secure and Personalized Medical Question Answering. In *Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, September 03, 2023. ACM, Houston TX USA, 1–6. https://doi.org/10.1145/3584371.3613016

[46] Niroop Channa Rajashekar, Yeo Eun Shin, Yuan Pu, Sunny Chung, Kisung You, Mauro Giuffre, Colleen E. Chan, Theo Saarinen, Allen Hsiao, Jasjeet Sekhon, Ambrose H. Wong, Leigh V. Evans, Rene F. Kizilcec, Loren Laine, Terika Mccall, and Dennis Shung. 2024. Human-Algorithmic Interaction Using a Large Language Model-Augmented Artificial Intelligence Clinical Decision Support System. May 2024. Association for Computing Machinery (ACM), 1–20. https://doi.org/10.1145/3613904.3642024

[47] Koustuv Saha and Amit Sharma. 2020. Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. *Proceedings of the International AAAI Conference on Web and Social Media* 14, (May 2020), 590–601. https://doi.org/10.1609/icwsm.v14i1.7326

[48] Joni Salminen, Chang Liu, Wenjing Pian, Jianxing Chi, Essi Häyhänen, and Bernard J. Jansen. 2024. Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions. May 2024. Association for Computing Machinery (ACM), 1–20. https://doi.org/10.1145/3613904.3642036

[49] Akio Sashima, Mitsuru Kawamoto, Satoshi Yazawa, and Kazuo Hiraki. 2024. Demo: CANSASI: Mobile Sensing Platform powered by Large Language Models. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*, February 28, 2024. ACM, San Diego CA USA, 152–152. https://doi.org/10.1145/3638550.3643047

[50] Burcu Sayin, Pasquale Minervini, Jacopo Staiano, and Andrea Passerini. 2024. Can LLMs Correct Physicians, Yet? Investigating Effective Interaction Methods in the Medical Domain. (March 2024). Retrieved from http://arxiv.org/abs/2403.20288

[51] Woosuk Seo, Sun Young Park, Mark S. Ackerman, Chan-Mo Yang, and Young-Ho Kim. Towards Designing a Safe and Reliable LLM-driven Chatbot for Children. *Proceedings of (1st HEAL Workshop at CHI Conference on Human Factors in Computing Systems) 1*.

[52] Woosuk Seo, Chanmo Yang, and Young-Ho Kim. 2024. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. 1–20. https://doi.org/10.1145/3613904.3642152

[53] William Seymour, Max Van Kleek, Reuben Binns, and Dave Murray-Rust. 2022. Respect as a Lens for the Design of AI Systems. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, July 26, 2022. ACM, Oxford United Kingdom, 641–652. https://doi.org/10.1145/3514094.3534186

[54] SIGCHI. ACM SIGCHI. Retrieved June 27, 2024 from https://sigchi.org/

[55] Frank Sommers, Alisa Kongthon, and Sarawoot Kongyoung. 2024. Fine-Tuning Large Language Models for Private Document Retrieval: A Tutorial. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, May 30, 2024. ACM, Phuket Thailand, 1319–1320. https://doi.org/10.1145/3652583.3658419

[56] Ritesh S Soun and Aadya Nair. 2023. ChatGPT for Mental Health Applications: A study on biases. In *The Third International Conference on Artificial Intelligence and Machine Learning Systems*, October 25, 2023. ACM, Bangalore India, 1–5. https://doi.org/10.1145/3639856.3639894

[57] Logan Stapleton, Sunniva Liu, Cindy Liu, Irene Hong, Stevie Chancellor, Robert E. Kraut, and Haiyi Zhu. 2024. "If This Person is Suicidal, What Do I Do?": Designing Computational Approaches to Help Online Volunteers Respond to Suicidality. May 2024. Association for Computing Machinery (ACM), 1–21. https://doi.org/10.1145/3613904.3641922

[58] Ian Steenstra, Prasanth Murali, Rebecca B. Perkins, Natalie Joseph, Michael K Paasche-Orlow, and Timothy Bickmore. 2024. Engaging and Entertaining Adolescents in Health Education Using LLM-Generated Fantasy Narrative Games and Virtual Agents. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–8. https://doi.org/10.1145/3613905.3650983

[59] Xin Sun, Yunjie Liu, Jan De Wit, Jos A. Bosch, and Zhuying Li. 2024. Trust by Interface: How Different User Interfaces Shape Human Trust in Health Information from Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 02, 2024. ACM, Honolulu HI USA, 1–7. https://doi.org/10.1145/3613905.3650837

[60] Aishwarya Surani and Sanchari Das. 2022. Understanding Privacy and Security Postures of Healthcare Chatbots. (2022). https://doi.org/10.1145/3313831.XXXXXXX

[61] Florian Szczepaniak, Jérôme Boudy, Gérard Chollet, Mossaab Hariz, and Christophe Lohr. 2023. A virtual Assistant Dedicated to the Accompaniment of the Person Informed of his Life Context thanks to the Smartphone. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, July 05, 2023. ACM, Corfu Greece, 662–666. https://doi.org/10.1145/3594806.3596594

[62] Yilin Tang, Liuqing Chen, Ziyu Chen, Wenkai Chen, Yu Cai, Yao Du, Fan Yang, and Lingyun Sun. 2024. EmoEden: Applying Generative Artificial Intelligence to Emotional Learning for Children with High-Function Autism. May 2024. Association for Computing Machinery (ACM), 1–20. https://doi.org/10.1145/3613904.3642899

[63] CNLD Testing & Therapy. 2022. How Long Should a Child's Attention Span Be? *CNLD Testing & Therapy*. Retrieved June 27, 2024 from https://www.cnld.org/how-long-should-a-childs-attention-span-be/

[64] Andrea C. Tricco, Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah D. J. Peters, Tanya Horsley, Laura Weeks, Susanne Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garritty, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiser, Jo Moriarty, Tammy Clifford, Özge Tunçalp, and Sharon E. Straus. 2018. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 169, 7 (October 2018), 467–473. https://doi.org/10.7326/M18-0850

[65] Michele M. Tugade and Barbara L. Fredrickson. 2004. Resilient Individuals Use Positive Emotions to Bounce Back From Negative Emotional Experiences. *J Pers Soc Psychol* 86, 2 (February 2004), 320–333. https://doi.org/10.1037/0022-3514.86.2.320

[66] Stanislav Ustymenko and Abhishek Phadke. 2024. Promise and Challenges of Generative AI in Healthcare Information Systems. In *Proceedings of the 2024 ACM Southeast Conference, ACMSE 2024*, April 2024. Association for Computing Machinery, Inc, 223–228. https://doi.org/10.1145/3603287.3651196

[67] Jiayin Wang, Weizhi Ma, Peijie Sun, Min Zhang, and Jian-Yun Nie. 2024. Understanding User Experience in Large Language Model Interactions. Retrieved June 25, 2024 from http://arxiv.org/abs/2401.08329

[68] Xin Wang, Samer M Abubaker, Grace T Babalola, and Stephanie Tulk Jesso. 2024. Co-Designing an AI Chatbot to Improve Patient Experience in the Hospital: A human-centered design case study of a collaboration between a hospital, a university, and ChatGPT. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–10. https://doi.org/10.1145/3613905.3637149

[69] Zhiyuan Wang, Varun Reddy, Karen Ingersoll, Tabor Flickinger, and Laura E. Barnes. 2024. Rapport Matters: Enhancing HIV mHealth Communication through Linguistic Analysis and Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 02, 2024. ACM, Honolulu HI USA, 1–8. https://doi.org/10.1145/3613905.3651077

[70] Joseph Weizenbaum. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman & Co., USA.

[71] Ruolan Wu, Chun Yu, Xiaole Pan, Yujia Liu, Ningning Zhang, Yue Fu, Yuhan Wang, Zhi Zheng, Li Chen, Qiaolei Jiang, Xuhai Xu, and Yuanchun Shi. 2024. MindShift: Leveraging Large Language Models for Mental-States-Based Problematic Smartphone Use Intervention. May 2024. Association for Computing Machinery (ACM), 1–24. https://doi.org/10.1145/3613904.3642790

[72] Tao Xu, Fei Wang, Prithwish Chakraborty, Pei-Yun Sabrina Hsueh, Gregor Stiglic, Jiang Bian, Lixia Yao, Alexej Gossmann, and Florian Buettner. 2023. Workshop on Applied Data Science for Healthcare: Applications and New Frontiers of Generative Models for Healthcare. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, August 06, 2023. ACM, Long Beach CA USA, 5893–5894. https://doi.org/10.1145/3580305.3599225

[73] 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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1 (March 2024), 1–32. https://doi.org/10.1145/3643540

[74] Deepika Yadav, Prerna Malik, Kirti Dabas, and Pushpendra Singh. 2019. FeedPal: Understanding opportunities for chatbots in breastfeeding education of women in India. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (November 2019). https://doi.org/10.1145/3359272

[75] Kailai Yang, Tianlin Zhang, Ziyan Kuang, Qianqian Xie, Jimin Huang, and Sophia Ananiadou. 2024. MentaLLaMA: Interpretable Mental Health Analysis on Social Media with Large Language Models. In *Proceedings of the ACM on Web Conference 2024*, May 13, 2024. ACM, Singapore Singapore, 4489–4500. https://doi.org/10.1145/3589334.3648137

[76] 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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2 (May 2024), 1–35. https://doi.org/10.1145/3659625

[77] Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, Pratik Ghosh, Mercy Ranjit, Kenza Bouzid, Anton Schwaighofer, Fernando Pérez-García, Harshita Sharma, Ozan Oktay, Matthew Lungren, Javier Alvarez-Valle, Aditya Nori, and Anja Thieme. 2024. Multimodal Healthcare AI: Identifying and Designing Clinically Relevant Vision-Language Applications for Radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–22. https://doi.org/10.1145/3613904.3642013

[78] Yue You, Chun Hua Tsai, Yao Li, Fenglong Ma, Christopher Heron, and Xinning Gui. 2023. Beyond Self-diagnosis: How a Chatbot-based Symptom Checker Should Respond. *ACM Transactions on Computer-Human Interaction* 30, 4 (September 2023). https://doi.org/10.1145/3589959

[79] Jordyn Young, Laala M. Jawara, Diep N. Nguyen, Brian Daly, Jina Huh-Yoo, and Afsaneh Razi. 2024. The Role of AI in Peer Support for Young People: A Study of Preferences for Human- and AI-Generated Responses. May 2024. Association for Computing Machinery (ACM), 1–18. https://doi.org/10.1145/3613904.3642574

[80] Yanbing Zhu, Sitian Xu, Boyang Liu, and Yuntao Jia. 2023. Optimization of Smart Healthcare Services and Development Strategies Based on Large Language Models. In *Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering*, November 17, 2023. ACM, Dalian China, 1004–1011. https://doi.org/10.1145/3652628.3652794

[81] Yiye Zhu, Jianping Shen, and Yinsheng Li. 2014. A human-centric user model for intelligent healthcare. In *Proceedings - 11th IEEE International Conference on E-Business Engineering, ICEBE 2014 - Including 10th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC 2014 and 1st Workshop on E-Commerce Engineering, ECE 2014*, December 2014. Institute of Electrical and Electronics Engineers Inc., 13–18. https://doi.org/10.1109/ICEBE.2014.15

[82] Anne Zimmerman, Joel Janhonen, and Emily Beer. 2023. Human/AI relationships: challenges, downsides, and impacts on human/human relationships. *AI and Ethics* (October 2023). https://doi.org/10.1007/s43681-023-00348-8

[83] **宋淑超以人**为本的可解释智能医疗综述, Song Shuchao, Chen Yiqiang, Yu Hanchao, Zhang Yingwei, and Yang Xiaodong. **第** 3***卷 第***期** 计算 **机** 辅助 设 计与 图 **形学 学**报 Review of Human-centered Explainable AI in Healthcare. https://doi.org/10.3724/SP.J.1089.202*.2024-00052

[84] OpenAI. Retrieved June 26, 2024 from https://openai.com/

[85] Making AI helpful for everyone - Google AI. Retrieved June 26, 2024 from https://ai.google/

[86] Meta Llama. Retrieved June 26, 2024 from https://llama.meta.com/

[87] Scoping. *PRISMA statement*. Retrieved June 27, 2024 from https://www.prisma-statement.org/scoping

[88] Conference Proceedings – ACM SIGCHI. Retrieved June 27, 2024 from https://archive.sigchi.org/conferences/conference-proceedings/

[89] ArXiv.org | Rutgers University Libraries. Retrieved June 27, 2024 from https://www.libraries.rutgers.edu/databases/arxiv-org

[90] Quantifying attention span across the lifespan - PMC. Retrieved June 27, 2024 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10621754/

[91] A/B Testing. *Emarsys*. Retrieved June 27, 2024 from https://emarsys.com/learn/glossary/a-b-testing/