Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults

ZIQI YANG*, University of Michigan, USA
XUHAI XU*, Massachusetts Institute of Technology, USA
BINGSHENG YAO, Rensselaer Polytechnic Institute, USA
ETHAN ROGERS, Northeastern University, USA
SHAO ZHANG, Northeastern University, USA
STEPHEN INTILLE, Northeastern University, USA
NAWAR SHARA, MedStar Health Research Institute, USA
GUODONG (GORDON) GAO, John Hopkins University, USA
DAKUO WANG†, Northeastern University, USA

Despite the plethora of telehealth applications to assist home-based older adults and healthcare providers, basic messaging and phone calls are still the most common communication methods, which suffer from limited availability, information loss, and process inefficiencies. One promising solution to facilitate patient-provider communication is to leverage large language models (LLMs) with their powerful natural conversation and summarization capability. However, there is a limited understanding of LLMs' role during the communication. We first conducted two interview studies with both older adults (N=10) and healthcare providers (N=9) to understand their needs and opportunities for LLMs in patient-provider asynchronous communication. Based on the insights, we built an LLM-powered communication system, Talk2Care, and designed interactive components for both groups: (1) For older adults, we leveraged the convenience and accessibility of voice assistants (VAs) and built an LLM-powered conversational interface for effective information collection. (2) For health providers, we built an LLM-based dashboard to summarize and present important health information based on older adults' conversations with the VA. We further conducted two user studies with older adults and providers to evaluate the usability of the system. The results showed that Talk2Care could facilitate the communication process, enrich the health information collected from older adults, and considerably save providers' efforts and time. We envision our work as an initial exploration of LLMs' capability in the intersection of healthcare and interpersonal communication.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; • **Applied computing → Health informatics**.

Additional Key Words and Phrases: Older adults, Patient-provider communication, Large-language-model

*Both authors contributed equally to this research.
†Corresponding author d.wang@northeastern.edu

Authors' Contact Information: Ziqi Yang, yangziqi@umich.edu, yangziqi@umich.edu, University of Michigan, USA; Xuhai Xu, xoxu@mit.edu, Massachusetts Institute of Technology, USA; Bingsheng Yao, yaob@rpi.edu, Rensselaer Polytechnic Institute, USA; Ethan Rogers, rogers.eth@northeastern.edu, Northeastern University, USA; Shao Zhang, zhang.shao1@northeastern.edu, Northeastern University, USA; Stephen Intille, s.intille@northeastern.edu, Northeastern University, USA; Nawar Shara, Nawar.Shara@Medstar.net, MedStar Health Research Institute, USA; Guodong (Gordon) Gao, gordon.gao@jhu.edu, John Hopkins University, USA; Dakuo Wang, d.wang@northeastern.edu, Northeastern University, USA.

ACM 2474-9567/2024/5-ART73
https://doi.org/10.1145/3659625

ACM Reference Format:
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, Article 73 (June 2024), 35 pages. https://doi.org/10.1145/3659625

1 INTRODUCTION

The healthcare needs of older adults are uniquely critical, with a significant proportion living with chronic diseases that necessitate long-term monitoring and regular interaction with healthcare providers [5]. According to the National Council on Aging (NCOA), nearly 95% of older adults in the US have at least one chronic health condition, and about 80% have multiple conditions [5], amounting to over 50 million seniors needing healthcare support. On the one hand, older people often experience a gradual decline in cognitive and physical abilities, which can make regular outpatient visits increasingly challenging. On the other hand, healthcare providers are already overloaded and have limited bandwidth [112, 115]. Asking them to visit patients' homes to check their health is not a scalable solution due to the significant workload [107]. As such, accessible home-based healthcare, which bridges the gap between older adults and their healthcare providers, emerges as one of the promising solutions. It can offer a familiar and convenient environment to older adults while saving healthcare providers' efforts.

Research has focused on various home-based telehealth applications to support asynchronous communication between older adults and their healthcare providers, such as remote health monitoring [15, 34, 47, 59, 77], healthcare plan management [11], and patient-provider collaborative review [30, 31, 98]. However, these platforms only support specific data formats and contexts that are tailored to some specific common diseases. When it comes to various daily health communication scenarios (*e.g.*, unexpected health concerns), older adults often have to refer to traditional asynchronous communication methods such as a basic text or voice messaging system, as healthcare providers are often not available in real-time for a call (*e.g.*, with a busy schedule or outside working hours). However, there are critical gaps in these methods. Firstly, getting feedback and progress via asynchronous messages is time-consuming, and digital messages could be lost due to providers' hasty schedules [69]. Also, older adults may fail to identify crucial symptoms in self-report messages due to limited health literacy. Although questionnaires can be used for more informative communication, they are mostly fixed and lack adaptability for different individual conditions [65, 89]. Meanwhile, older adults generally have lower technology proficiency compared to younger generations, which is another common obstacle to accessing and using some of these tools [40, 45, 65, 104]. For healthcare providers with certain patient follow-up protocols (*e.g.*, post-surgery follow-up), making manual phone calls is one of the most common practices, yet it costs a significant amount of workload, and sometimes older adults may not be reachable. The issues in efficiency, adaptability, and accessibility within current communication methods exert communication challenges on older adults and their healthcare providers.

A novel conversational interface between older adults at home and healthcare providers may fill in these gaps in asynchronous communication by providing an accessible, interactive, and intelligent communication system. The recent technological boost of large language models (LLMs) brings us new opportunities to better achieve the goal. The HCI community has explored the application of LLMs for various healthcare practices, such as mental health support [66, 68], health information seeking [81, 117, 128], and public health interventions [58], revealing LLMs' intelligence in processing health-related information. However, there is little previous work leveraging LLMs for patient-provider communication. In our case, we can leverage its outstanding power of natural language conversation generation and information summarization. For older adults, an LLM-based conversational agent (CA) may provide convenience and easy accessibility. Given older adults' limited health literacy and complicated health conditions, the LLM-powered CA may help older adults better convey their information asynchronously

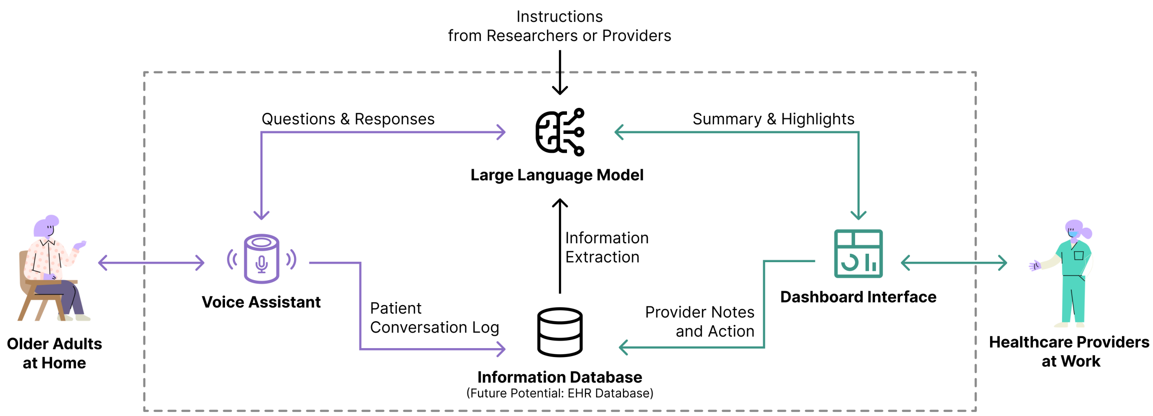

Fig. 1. Overview of Talk2Care System. The system consists of two modules. 1) The patient module: An LLM-powered VA interface (in purple) that generates natural conversation with home-based older adults to collect health information and forward it to healthcare providers; 2) The provider module: A dashboard interface (in green) that summarizes the key information from the older patient conversation to assist providers who are responsible for communication (*e.g.*, nurses and patient navigators). Note that Talk2Care does not provide specific healthcare advice. Our current implementation does not involve an actual electronic health record (EHR) system, which can be a promising future direction.

to providers when needed, such as in cases with unexpected health concerns. Meanwhile, providers can also use LLMs to automate the information-gathering and summarization process with high-quality data collected from older adults.

However, the role of LLMs in the asynchronous communication between older adults and healthcare providers is still unclear. To address this gap, we first conducted two semi-structured interviews to understand communication challenges and older adults' and providers' perceptions towards LLMs-based communication systems. Our studies involved both older adults with various health conditions (N=10), as well as health providers (N=9). Our results revealed a set of asynchronous communication challenges between older adults and healthcare providers. We found that it is hard for patients and providers to reach each other, and much information often requires follow-up, explanation, and analysis. These results illustrate the high potential of an LLM-based communication tool.

Leveraging the insights from our interview study, we designed and implemented a pilot LLM-based system, Talk2Care, to support diverse home-based healthcare communication scenarios. The system consists of two modules. (1) A patient-facing module: A conversational interface presented as a voice assistant (VA) that provides a natural conversation experience with older patients to collect health information. (2) A provider-facing module: A dashboard interface for health providers that summarizes and highlights key information from patient-VA conversations. It is worth noting that our system *does not* aim to provide any specific health advice due to the early-stage exploratory nature of this work and ethical concerns. As discussed later, the ethical concerns must eventually be addressed, but in this work, we explore the value of LLM-driven communication as a first step. We evaluated the pilot Talk2Care in two cases close to real-world scenarios, one usually initiated by older adults (unexpected health concerns) and one usually initiated by providers (post-surgery follow-up). We conducted two user studies with both sides to evaluate the usability of Talk2Care. Our results suggest that our LLM-powered communication tool can facilitate the asynchronous communication process, enrich the health information collected from older adults, and save providers' efforts and time considerably.

Our primary contributions are:

- We conducted two interview studies with both older adults and healthcare providers to understand the communication challenges and opportunities for LLMs to assist patient-provider communication.
- We designed and developed a pilot LLM-powered system, Talk2Care – with a VA interface for older patients and a dashboard interface for providers – to facilitate home-based patient-provider communication for older adults.
- We evaluated our systems with two common home-based healthcare scenarios via two user studies with older adults and providers to demonstrate the usability and effectiveness of Talk2Care.
- We revealed potentials and design implications for future LLM-powered systems to facilitate and mediate remote patient-provider communication in a human-AI collaborative workflow.

2 RELATED WORK

We first provide an overview of telehealth and patient-provider communication solutions (Section 2.1). We then briefly overview intelligent voice assistants for older adults' healthcare (Section 2.2) and large language models for healthcare (Section 2.3).

2.1 Telehealth and Patient-Provider Communication Solutions

Telehealth, or telemedicine, refers to the use of digital communication technologies to provide healthcare services remotely [109]. In the past decade, research has shown the advantages of telehealth in saving patients' and providers' time and effort compared to in-person visits [23, 28, 34, 38], especially during the pandemic isolation [19]. Researchers have explored novel and more accessible tools to enrich the amount of patient health information in presented artifacts for patient-provider collaborative review and remote patient management, such as mobile health applications [32, 36, 71, 74, 114, 122, 126], sensor systems and instruction platforms [20, 43, 47, 62, 88, 123, 125], personal coaches [11], online platforms for healthcare consultation [110, 111]. However, since these applications are usually designed for a specific common disease or developed in an isolated platform with a specific format, they are not adaptable to different patient needs, hard to integrate with existing healthcare systems, and has poor usability [37, 38, 65, 89, 101, 112]. Especially for older adults who have relatively low health and technical literacy, most of the existing solutions still cannot provide adequate accessibility and usability [40, 65, 92, 104]. Therefore, a great number of older patients, still refer to traditional communication methods such as phone calls [38, 92]. Similarly, healthcare providers also often find these telehealth tools hard to use [57, 60, 61]. Therefore, making phone calls and leaving messages are still popular communication methods for providers in cases such as hospital discharge follow-up [60, 78]. However, these solutions lacked efficiency and adaptivity. Due to limited provider availability, communicating via asynchronous messages is time-consuming, and digital messages could be lost due to providers' hasty schedules [69]. Also, when the messages rely on older adults' descriptive self-reports, crucial symptoms could be missed due to their limited literacy [29, 96]. Another option is to integrate structured questionnaires, yet they are not adaptive or flexible enough for individuals' personalized conditions and needs [39, 75]. These limitations in current communication technologies call for an intelligent and accessible solution to facilitate their patient-provider. In this paper, we explore a novel, easy-to-use, and natural communication solution for both home-based older adults and healthcare providers.

2.2 Intelligent Conversational Agents for Older Adults Healthcare

Researchers have explored a variety of CAs to effectively support older adults' home-based healthcare needs, including chatbots [44, 97], embodied or physical CAs [13, 14, 103, 105], and intelligent VAs [17, 35, 49, 91]. Among them, VAs have stood out with their simplicity, convenience, and good accessibility in scenarios including health information seeking [17, 21, 49, 56, 91], learning [35], assisting home health aides [13, 14], and mental support [103, 105, 106]. Yet studies have also pointed out the limited use of these VAs by older adults due to usability barriers, cultural gaps, lack of explainability, and flexibility [18, 44, 49, 108, 131], and none of them have explored CAs as communication interfaces. This motivated us to explore novel CAs to fill in the gaps in older adults' home-based healthcare and communication, taking VA as an example.

Some recent studies have explored the combination of conversational user interfaces and more intelligent techniques. For example, Natural Language Processing (NLP) methods have been used to build CAs in travel apps and browsers to simplify user interaction [25, 90], and LLMs have been used in CAs to interact with mobile UI and programming tasks [94, 113]. Recently, Zhang *et al.* proposed SpeechGPT, an LLM with intrinsic cross-modal conversational abilities [129]. Rubenstein *et al.* presented AudioPaLM, a unified speech-text LLM [95]. However, little is known about whether these could be applied to real-world contexts, especially for home-based older adults. Our work seeks to leverage this interaction paradigm to build an interactive and easy-to-use LLM-powered CAs that advances personalized information gathering and asynchronous communication with healthcare providers.

2.3 Large Language Models for Healthcare

Recently, the technology boost of Large Language Models (LLM), such as ChatGPT [87], has shown its great potential in engaging in, scaffolding, and processing natural conversations [54, 72, 102, 113, 120, 121]. The latest studies have discussed and explored its application for individual patients' in health information seeking [119, 128], mental health support [66, 68, 124], personal health coaching and management [81, 117] and health education [67, 81]. For healthcare professionals, researchers have also leveraged LLMs to support public health interventions [58], clinical pre-screening [48, 115], risk prediction [16, 42, 63, 86] and information processing [9, 10, 26, 64, 84, 85]. However, multiple studies have raised concerns about using LLMs for healthcare scenarios, due to their inconsistency, potential errors, and bias in clinical work [17, 55, 64]. Due to these concerns, in our work, we focus on facilitating communication, but not on providing any specific healthcare advice.

Despite the great potential of LLM for promoting healthcare work and experiences, little work aims to address the challenges of patient-provider communication mentioned in Section 2.1, and little is know about how LLM might help patient-provider communication and benefit both stakeholders. Close to our research, Jo *et al.* conducted interviews with stakeholders in an LLM-powered system that supports public mental health interventions for people living alone [58]. The study revealed that the system could reduce the workload for mental health practitioners while providing mental support for end users in large-scale healthcare services. However, the existing solution still failed to meet the personalized requirements of users at home because it followed a uniform setup("one-size-fits-all"), which lacked the necessary flexibility for customization [58]. With our pilot system, we aim to facilitate patient-provider communication by providing older adults with more interactive, responsive, and adaptive experience, and providing healthcare providers with more efficient information analysis assistance.

3 NEED-FINDING STUDY

To better understand the communication gap between older adults and their healthcare providers and to identify user needs, we conducted two semi-structured interview studies, one with older adults (Section 3.1), and the other with healthcare providers (Section 3.2). Through our interview studies, we identify (1) communication challenges such as limited provider availability and disparities in literacy, (2) the opportunities for future LLM-powered systems to offload providers and provide mental support, as well as (3) reliability and ethical concerns. Understanding these challenges and user needs suggests the potential of LLM-powered systems to facilitate communication. Moreover, our results guide the design of our system in Section 4.

3.1 Interview Study 1: Older Adults

3.1.1 Methods: Participants and Procedure. After obtaining the IRB approval, we recruited older adults (N=10) from a local university-maintained participant pool of older adults with snowball sampling. Participants lived in an area in the northeast part of the United States. All the older adults resided in single-family homes and were 65 years of age or older. Participants reported one or more chronic conditions and urgent care histories; they had rich experience communicating with their healthcare providers in recent years. Out of the ten older adult participants, two of them had experience using AI (*e.g.*, using ChatGPT on its web interface), six knew about the concept but had not used AI, and two others were unfamiliar with the concept. Table 1 summarizes their detailed demographic information[1].

Table 1. Demographics of Older Adult Participants

P#	Age	Gender	Chronic/Occasional Condition	Surgery/Accident History
P1	65-69	Female	Migraine, hypertension, former depression, dementia risk	Joint surgery
P2	75-79	Female	Hypertension, arthritis, stomach issue, anxiety disorder	Thumb injury
P3	65-69	Female	Generally healthy, sinus infection, cervical dystonia	/
P4	80-84	Female	Generally healthy, former alopecia	Ankle injury
P5	65-69	Female	Generally healthy	Throat & shoulder surgery, body injury
P6	80-84	Female	Atrial fibrillation, occasional pain	/
P7	70-74	Female	Generally healthy, usual aches and pain	/
P8	70-74	Male	COPD, sleep apnea, atrial fibrillation	/
P9	65-69	Male	Generally healthy, long-term COVID influence	Collarbone injury
P10	70-74	Female	Arthritics, occasional depression, pain and hypertension	/

Our interview sessions followed a semi-structured format. Our key interview questions covered three perspectives: (1) What were older adults' needs and challenges in communicating with their healthcare providers at home? (2) How might an LLM-based CA system facilitate communication and address these needs and pain points? (3) What concerns and risks would they have when applying such an AI-powered system? For the participants who had no experience interacting with LLMs, we used slides to introduce the concept of AI and LLM in descriptive words and visual cues and presented an example conversation with ChatGPT on the web interface. Based on these starting points, the interviewer followed up with interviewees based on their responses. The specific questions can be found in Appendix A.1 and Appendix A.2. The study lasted 25-30 minutes. Participants were compensated $25 for their time. The interviews were recorded via Zoom and transcribed afterward. Thematic analysis was employed to analyze the data. Two authors coded interview transcripts iteratively until they reached a consensus. We summarize our findings as follows.

3.1.2 Findings: Communication Needs and Challenges. Participants reported a variety of examples where they needed to communicate directly with their healthcare providers at home, such as daily questions about new symptoms and urgent care (*e.g.*, COVID or flu symptoms), confusion or changes about an existing condition (*e.g.*, occasional pain or blood pressure change, prescription instruction), and post-surgery-related questions. Many cases were consistent with, and supported by, prior research [37, 43, 80, 112]. Most of these communication needs were case-specific and personalized. Participants' current solutions included visiting/calling healthcare services,

[1] We show the age of participants by age range for deidentification purposes.

communicating through the patient portal with providers, discussing with friends or family, and seeking online resources as alternatives. Our interview revealed a set of challenges to these existing solutions.

Communication Inefficiency Despite the wish to get quick and in-time responses from the healthcare system and providers, 5 participants mentioned that it was **hard to get to healthcare providers** because they were often too busy during working hours. For example, " *then you're on hold for maybe 30 minutes, which is crazy.*" (P10) The in-person visiting option posed more challenges. Besides the traveling difficulties, the waiting time for an appointment with providers often took several weeks or even months.

Even in cases where participants got to talk to providers, some information was often ambiguous or even **lost during communication**. For example, P1 mentioned that nurse practitioners "*never put down exactly what I say... They may not realize that it is very significant*". P8 gave another example where the messages from providers and lab results often lacked explanations and were hard to understand. Such an information loss caused the participant to pay extra effort to describe things or request explanations one more time when they talked to doctors in subsequent interactions, which reduced communication efficiency and brought confusion for older adults.

Some participants were more familiar with technology and learned to use digital patient communication portals to leave messages to providers, which could address these challenges to some extent. However, **lack of digital literacy** is a widely known issue [40, 91, 104]. Only four out of the ten participants knew how to use such digital technology.

Reliability Concerns of Alternative Resource. Since they could not get in-time communication with health providers, four participants reported resorting to alternative resources by searching online or talking to families. Consistent with previous work [17, 21], they expressed concerns about the reliability of these resources. "*I'd rather just go to the doctor. That's their area of expertise.*" (P1) This became another strong motivation for us to facilitate communication between older adults and providers.

3.1.3 Findings: Opportunities for An LLM-facilitated CA System. To further understand how LLMs may help older adults with their communication, we introduced the concept of an LLM-based CA with the possibility of implementing a VA as the user interface. We also presented examples of how such a system could have natural conversations, collect information and conditions, and forward the information to healthcare providers for further decision and contact. We highlighted explicitly that our system would not provide any specific health advice, but the system could provide general information about clinics, procedures, or term explanations. Overall, older participants expressed strong interest in such an AI-powered system to help them address the communication challenges.

Offloading for Healthcare Providers. As introduced in Section 3.1.2, one of the key communication challenges was the unavailability of healthcare providers. Older adults hoped that the system could help save providers' efforts. "*... Whereas with a computer, and assuming it can respond to 100 people at the same time ...*" (P8) This could potentially help providers with tedious procedure work and give them the bandwidth to respond to patients' actual inquiries.

Emotional and Accessibility Support Although we made it clear that the system would not provide specific health advice, 3 participants still thought it could give immediate responses and provide **emotional support**. For example, P4 thought the system could accompany her at home since she lives alone. P10 also agreed that the audio modality of the VA would be comforting, "*Sometimes you just wanted to ... hear a voice*" (P10). Moreover, the support could also be expanded to other family members. For example, P1 shared her post-surgery experience and thought that if such a system existed, it would "*give my daughter enough comfort to let me be in our apartment alone after*" (P1).

In addition, participants commented that the natural VA system could address the technology barrier of navigating digital communication portals by improving cognitive accessibility the intelligence from LLMs. For example, P1 had a potential dementia risk and said "*there's going to come a time when I'm not going to be able to remember how to use it.*"

Function Expectations. In addition to the natural conversation, information summary, and forwarding features, participants also suggested other expectations, including having pre-set personalized instructions from doctors (P8), providing high-level suggestions and references (P1, P6), and delivering administrative information (P2, P10).

3.1.4 Findings: Concerns and Risks. We also investigated the potential concerns of older adults when using the intelligent system. We found the common ethical concerns for AI systems [130] and privacy concerns of VAs from 3 participants [91]. For example, P5 was concerned that they would accidentally reveal sensitive personal information: "*because it's a little bit [listening to] everything*". However, many participants did not have such concerns, as long as it "*follows the criteria that the federal regulations*" (P2). Besides, some participants were also concerned about the learning curve of the system, which we will evaluate through a user study in Section 5.2.

3.2 Interview Study 2: Healthcare Providers

3.2.1 Methods: Participants and Procedure. By emails and snowball sampling, we recruited 9 nurse practitioners, doctors who are primary care physicians or specialists, and other professionals who have close contact with patients. Participants have frequent communication with patients at home, and most of them have close communication with older adult patients. Six of the nine participants had experience with AI (*e.g.*, using ChatGPT on its web interface), and three others knew about the concept. Table 2 summarizes their demographics and expertise. We mainly recruited physicians because we focus on the communication needs concerning clinical decisions or action in this study. In these scenarios crucial to older adults' health, physicians are the ones who are mainly responsible, while nurse practitioners or other staff largely worked on administrative work. In addition, the older adult participants expressed their strong wish to closely communicate with physicians, as discussed in Section 3.1.2 We include providers with different roles to gain a holistic view of the providers' perspectives.

Table 2. Demographics of Healthcare Provider Participants

PP#	Gender	Role, Responsibility, Specialty	Years of Experience in Healthcare
PP1	Female	Primary care physician, specialist in diabetes	10-20 Years
PP2	Female	Nurse practitioner in emergency department	0-2 Years
PP3	Female	Dentist, health consultant	0-2 Years
PP4	Male	Palliative medicine physician	>20 Years
PP5	Male	Primary care physician, internist, pediatrician, hospice and palliative care specialist	>20 Years
PP6	Male	Internal medicine physician	2-5 Years
PP7	Female	Clinical research assistant & medical records coordinator	5-10 Years
PP8	Female	Palliative care nurse practitioner, former patient navigator	>20 Years
PP9	Female	Pediatrician, physician in ER	>20 Years

We designed a semi-structured interview process, with a focus on three perspectives: (1) What were healthcare providers' needs and challenges in communicating with their patients at home, especially with older adults? (2) How might an LLM-based system facilitate their communication with patients and address these needs and challenges? (3) What concerns and risks would they have when applying such a system? The specific questions can be found in Appendix A.2. The interviewer followed up with the providers based on their responses. The study lasted 25-30 minutes. Providers were compensated $25 for their time. Similar to our procedure in Section 3.1, we recorded the Zoom interview, and two researchers used thematic analysis on the transcribed data. We summarize our findings in the following sections from the three perspectives listed above.

3.2.2 Findings: Communication Needs and Challenges. Providers mentioned a variety of scenarios that required patient-provider communication. The majority were about older patients' daily care inquiries and the follow-up of previous treatments or surgeries. Sometimes long-term patient education also requires communication. Providers did use digital tools to manage information and reminders, yet for older adults, phone calls were still the most common method. Our interview results suggest several prominent challenges.

Ineffective Communication with Existing Asynchronous Methods. In Section 3.1.2, participants mentioned that providers were hard to reach. However, interestingly, providers felt the same. Four providers found patients, especially older adults, were hard to reach via phone calls. They often needed to call multiple times to finally reach patients, and sometimes they just could not succeed. Meanwhile, providers mentioned that they are "understaffed and overworked" since they usually have too many patients to handle (PP8). Their bandwidth was limited, so they could not spend all their time reaching out to older patients or going through a standardized in-person visit. As a result, providers had to prioritize other tasks, and communication often experienced inevitable delays.

Existing asynchronous solutions have been trying to address this limited availability challenge. Tools such as patient portals or EHR systems are set up as an alternative to manual phone calls. However, providers commented that the messages from patients were usually **unclear**. These messages were usually too limited or overly long for providers to make the right decision. Meanwhile, the design of current EHR systems has too much redundant information for providers to process "*There's so much information that you need to disregard.*" (PP5) This may explain the information loss issue mentioned by older adults in Section 3.1.2.

Extra Effort Communicating with Older Adults. Out of the nine providers, seven also mentioned the challenges during the real-time communication with older adults: Older adults often had limited health literacy and higher accessibility needs. Four of the providers felt that they needed to pay extra attention to language and spend more time explaining concepts to make sure that the information was delivered to and understood by older adults. "*... like talking [in a] slower pace or using non-medical terms that any patient can understand.*" (PP2)

3.2.3 Findings: Opportunities for An LLM-facilitated Communication System. To investigate how LLM-based AI can be used to help providers address communication challenges, we introduce the concept of LLM and our system. We briefly introduced the concept of both the patient-facing VA system for older adults and the LLM-powered information dashboard that summarizes the health information of older adults. Similar to Study 1, we also explicitly mentioned that the AI system would not provide any specific health advice. Overall, providers were excited about the potential of the system after learning the concept. We summarize the opportunities below.

Promoting Efficiency In Section 3.1.3, older adults expressed the hope that such an LLM-powered system could help offload providers. This was confirmed by our Study 2. Providers agreed that using a VA system & an information dashboard could significantly **save their effort**. "*[The system] could be very helpful for those local community clinics that may cover tens or even hundreds of patients.*" (PP1) Providers could leverage such an intelligent communication system to reduce the procedural process and focus on more critical tasks.

Moreover, providers also mentioned that such a system could help to **address the information loss problem** (Section 3.1.2 and 3.2.2) by recording things and provide analysis results. "*Sometimes we weren't able to capture the full experience... AI will come in if it's being recorded... and capture that data.*" (PP7) Providers would like to see a concise summary of the conversation as they saw during our study (PP1, PP2, PP3), which could help them with the triage and prioritization of the right next step. These comments guided the design of our dashboard system in Section 4.2.

Experience Improvement for Older Adults. Consistent with the comments from older adults, providers, from their own perspective, also agreed that the VA system could help with the conversation experience for older adults. "*[The VA system] might eliminate that feeling of judgment, and the patient will probably answer more accurately, as opposed to sitting across from another person who could potentially, you know, unintentionally making them feel uncomfortable because of the certain questions and answers that are required.*" (PP7) Providers also agreed with the accessibility advantage of a VA system over text-based solutions , which strengthens our motivation for the VA system for older adults.

Function Expectation. Providers also brought up a few features that the communication system would be good to have. For the VA system, several providers mentioned the necessity to provide "closed loop communication". "*That would be helpful and increase the accuracy of the information.*" (PP2) Meanwhile, providers also mentioned the expectation of such an AI system to be integrated with EHR systems (PP1, PP4). Some providers also brought up the importance of recording conversation audio to assist with certain diagnoses.

3.2.4 Findings: Concerns and Risks. Providers had a strong agreement with our design of not providing health advice due to the concern about AI reliability. Moreover, they also expressed concerns about the risks of misinterpretation (PP2, PP5), AI-embedded bias (PP3, PP7), and sensitive/triggering topics, such as cancers or sex crimes (PP4). Meanwhile, some were also worried about how AI might influence patient-provider interaction in the long run (PP4, PP9). Some of these ethical concerns were beyond the scope of this paper, and we will discuss potential solutions in Section 6.

3.3 Findings Summary

We summarize our key findings of this section as follows:

(1) Problems in current communication methods, including provider availability, limited information communicated, and accessibility needs, have led to communication challenges for older adults and their providers.
(2) Participants are highly interested in an LLM-powered system to offload healthcare providers, provide mental support, and gather more information.
(3) Reliability and privacy concerns were raised by both patients and providers.

These insights can guide the design of our LLM-powered system in the next section to facilitate patient-provider communication for older adults.

4 TALK2CARE SYSTEM

The insights from older patients and health providers shed light on the design and implementation of our pilot Talk2Care system to facilitate patient-provider communication. Figure 1 presents an overview of our pilot system. Talk2Care consists of two parts: First, the patient component: an LLM-powered VA interface that asks and answers older patients' questions to collect their health information, which will be forwarded to healthcare providers for future decisions and contact (Section 4.1). Second, the provider component: a dashboard interface that employs LLM to summarize patients-VA conversations and present important information while keeping original details to save effort and time for healthcare providers (Section 4.2).

4.1 Patient Component: LLM-powered Voice Assistant

The results in Section 3.1.3 confirmed the advantage of an LLM-powered VA system for its convenience and accessibility for home-based older patients. We build a VA interface to collect information health information (*e.g.*, detailed symptom descriptions). The interface has a multi-turn conversation ability so that they can have continuous and natural back-and-forth conversations with older patients, generating appropriate and personalized follow-up questions to collect important information. Figure 2 provides an overview of this component of

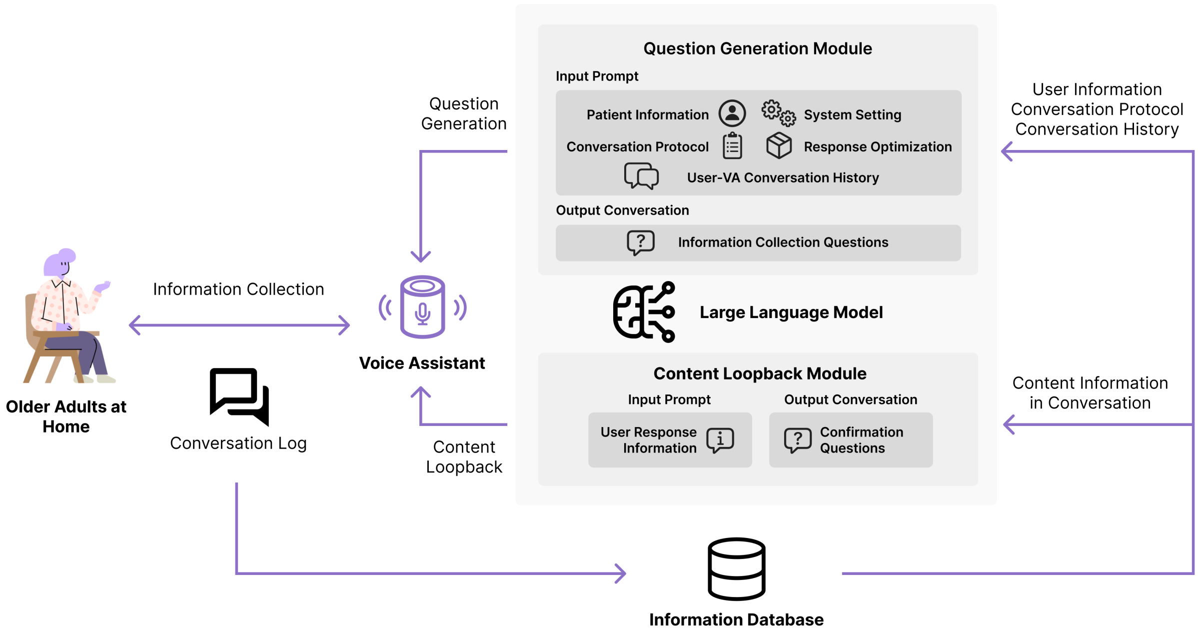

Fig. 2. The Component of Talk2Care System for Home-based Older Adults. The VA interface has multi-turn personalized conversations with the older adult to collect related health information. The LLM-powered Question Generation Module is responsible for taking the older adult's words and generating questions for effective information collection. The prompt design of this module is detailed in Figure 3. Another LLM-powered Content Loopback Module is to make sure that key information from the older adult (e.g., pain level) is accurate by double-checking the content, a common healthcare communication practice. The older adult's information, conversation protocol, and conversation log are stored in the information database.

Talk2Care system, including the two modules for conversation generation (Section 4.1.1 and 4.1.2), the VA hardware (Section 4.1.3), and the database (Section 4.1.4).

4.1.1 Question Generation Module. The core part of the VA system is to generate high-quality questions to collect key health information from older adults, which can then be forwarded to healthcare providers for decision-making. To achieve this goal, we leverage GPT-3.5-Turbo over GPT-4.0 due to the fast speed of GPT-3.5. LLMs are known to generate offensive or biased texts, sometimes even harmful content [51, 130]. To address the ethical concerns, we conduct a series of prompt designs to improve the question generation referring to prompt optimization practices from previous works [8, 117]. Specifically, one researcher conducted a set of preliminary prompt designs and tested them with GPT-3.5. Then, another two researchers, one with technical expertise and the other with healthcare expertise, acted as quality checkers for the content and provided feedback. This process was iterated through multiple rounds until all researchers were satisfied with the generated content. [2]

Finally, we propose a set of five important content factors to be integrated as a complete prompt, as illustrated in Figure 3. **(1) Patient Information.** This part includes the older adult's basic information, such as their name, gender, age, and living situation. Moreover, it also includes their health situations such as chronic conditions and medical history. This information was added to the database during the initial setup of the system. **(2) Conversation Protocol.** To ensure the validity of the information collection questions, the VA should follow a

[2]Despite these efforts, we acknowledge that this could not address the ethical concerns completely. We show more discussion in Section 6.

standard protocol employed by providers for a specific task (e.g., daily healthcare needs in Figure 3). Different scenarios may require different protocols (e.g., post-survey follow-up vs. daily care needs). Therefore, the information database stores a set of protocols collected from online resources, researchers, and providers. We separate the task summary and question protocol into two chunks to provide a flexible mapping between tasks and protocols. [3] **(3) System Setting.** This part instructs the role of the VA (as a communication facilitator), its responsibilities (i.e., collecting health information but not giving specific health advice), and its communication style (i.e., concise, health-focused, and easy-to-understand, as suggested in Section 3.2.2). The specific wording of our prompts can be found in Figure 3. **(4) Conversation History.** This part includes the conversation history between the older adult and the VA, which can augment the VA's memory and personalize the conversation experience. **(5) Response Optimization.** To optimize the VA's response to the older adult, this part is designed to improve the quality of each response and question generated by GPT-3.5. For each conversation session, (1) - (4) are only used at the beginning of the generation, while (5) is added in the prompt across all conversation rounds.

4.1.2 Content Loopback Module. In addition to the question generation, our interviews with providers (Section 3.2) suggest the importance of content confirmation, especially for important health-related information. Therefore, we design a simple content loopback module. When a question asks about specific values (e.g., the pain level), the module loops the value back to users and asks them a second time to confirm.

4.1.3 Voice Assistant Hardware. We experimented with multiple audio devices, including Raspberry Pi and Arduino boards. Finally, we decided to deploy our system on the Amazon Alexa platform by wrapping it as an Alexa Skill, so that we can leverage the well-integrated ecosystem. We use Alexa Echo Dot as the hardware. Voice-to-text and text-to-voice service provided by Amazon Alexa is employed to transcribe the older adult's input into texts and convert LLM-generated texts into audio, bridging the text content and natural speech interaction.

4.1.4 Information Database. We employ a cloud-based database to store the older adult's information, conversational protocols, as well as conversation log. These contents are used to construct the prompt for both the question generation module (1, 2, and 4 in Section 4.1.1) and the content loopback module. The database is encrypted to protect each individual's privacy.

4.1.5 Conversation Flow. Depending on the scenarios, the conversation can be initiated either by older patients (*e.g.*, with specific health concerns and care needs) or by the VA (*e.g.*, health condition follow-up requests). In each session, the VA will lead the conversation to ask the patient about symptoms and iteratively gather detailed information to pass to healthcare providers following the prompt instructions. For example, in the post-surgery scenario, the VA will first ask the patient about the overall health conditions and any pain or discomfort. If the patient mentions any pain or discomfort, the VA will ask about details (with loopback confirmation), such as pain level, positions of discomfort, and any actions that may influence the symptoms. If the patient has any confusion about the information, the VA explains the concepts accordingly but does not provide direct medical advice.

4.1.6 Exception Handling. We designed the system and prompts to handle common exceptions in voice recognition and user input to minimize unsatisfactory user experience in abrupt failure or termination. We use Alexa skill re-prompts [1] and conversation history to handle unexpected pauses, so that users could have continuous conversations with the VA after pauses. If voice recognition has errors, users could interrupt the VA and start again.

Some users might seek immediate medical advice from the system, and according to prompts, the system generates responses such as "*I'm not a doctor, but it would be best to consult with your healthcare provider about*

[3]Later in our user study in Section 5, we pre-determined the protocol depending on the scenarios. Future work can explore automatic protocol selection depending on the task, as we will discuss in Section 6.

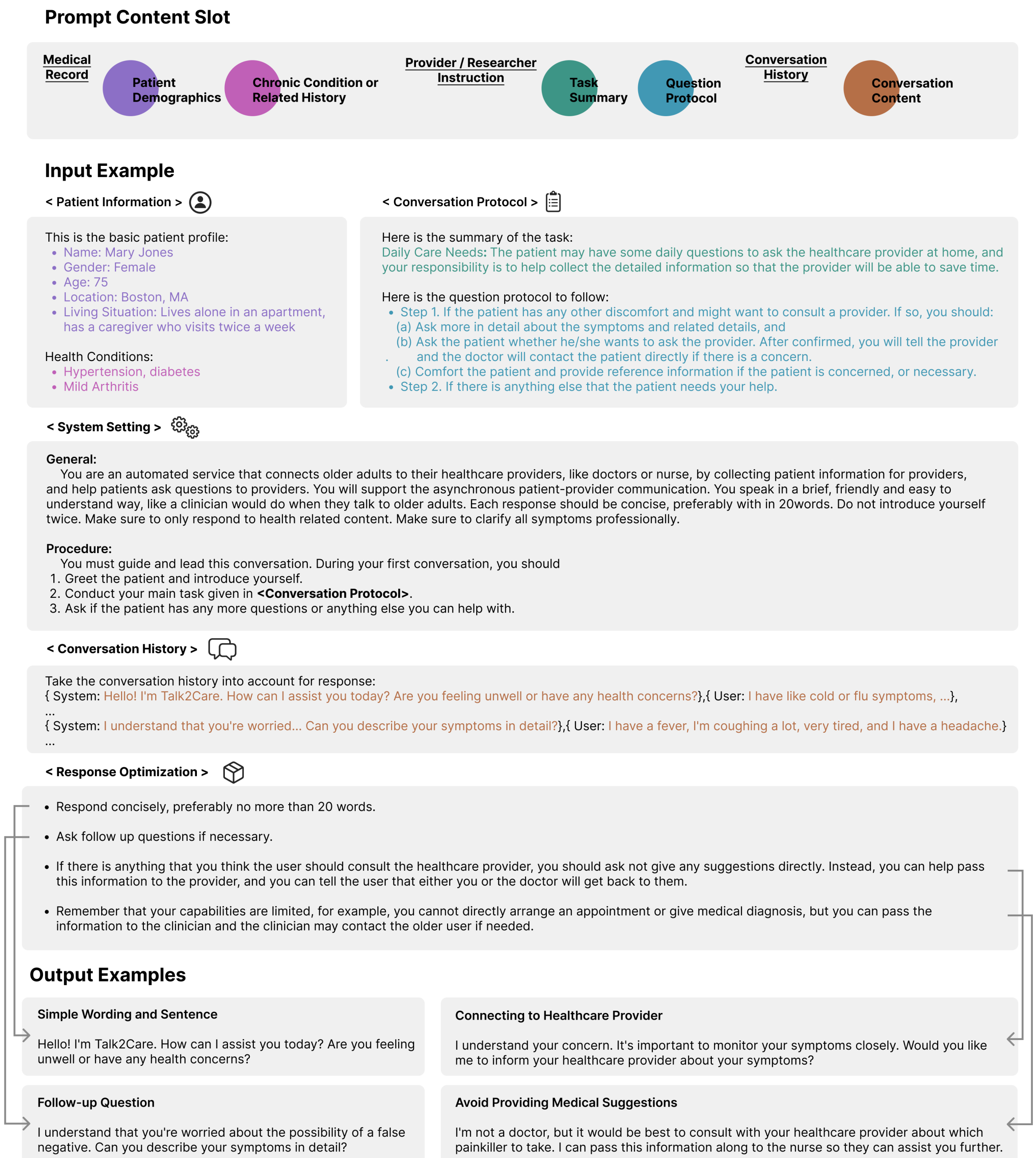

Fig. 3. Prompt Design of High-Quality Question Generation for Health Information Collection. The input prompt consists of five parts: 1) patient information, 2) conversation protocol, 3) system setting, 4) conversation history, and 5) response optimization. For multi-turn conversation, 5) will be repeated for each round of conversation. The colored texts are parameters that can be extracted from the information database (see Figure 2). Note that the conversation protocol needs to be set by researchers or healthcare providers to ensure question validity. This figure shows an example of daily-care protocol.

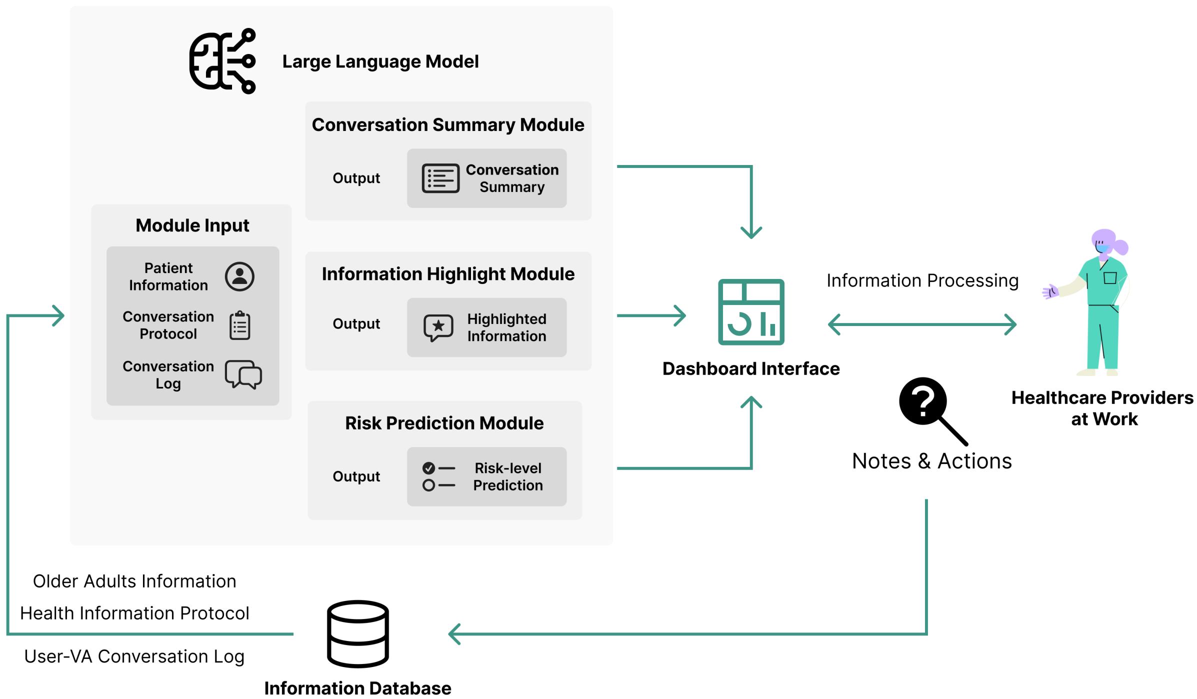

Fig. 4. The Component of Talk2Care System for Healthcare Providers. The information dashboard summarizes and highlights key older adults' information. The main content on the dashboard is generated by three LLM-powered modules: (1) The Content Summary Module formats the conversation log and user information into a clinical note structure. (2) The Information Highlight Module color-codes the parts in the conversation log that require attention. (3) The Risk Prediction Module suggests the health risk (low, moderate, and high) based on the current conversation log. Providers can take notes or further actions on the dashboard, which are then stored in the information database.

which painkiller to take. I can pass this information along to the nurse so they can assist you further. "(VA) Since we aim to take advantage of LLMs, we allow the LLM to provide reasonable explanations for general information assistance, such as "What is hospice and palliative care?" The strategies appeared to be successful in our user study in Section 5.2.

4.2 Provider Component: LLM-powered Information Dashboard

Other than the LLM-powered VA system for older adults, we further developed a dashboard interface for healthcare providers to review the information in the patient-VA conversation and take further actions if necessary. Figure 4 illustrates an overview of this component of Talk2Care, including three modules for information summary (Section 4.2.1), context highlight 4.2.2), and risk prediction (Section 4.2.3), together with the dashboard (Section 4.2.4) and the database (Section 4.2.5).

4.2.1 Conversation Summary Module. Our interview results with experts suggest that providing a summary of the conversation between the older adult and the VA system is very helpful (Section 3.2.3). This module aims to extract the key messages from the conversation log and format them into a clinical notes structure. To achieve this goal, we also leverage GPT-3.5 with a series of prompt designs.

By leveraging provider training tutorials and literature [2, 7, 8], our prompt iteration process is close to the one introduced in Section 4.1.1. The final prompt design structure, as shown in Figure 5, also resembles the question generation for older adults (Figure 3), including five factors. **(1) Patient Information.** This is the same information of the older adults from the information database. **(2) Conversation Protocol.** Other than providing the task summary and question protocol, the prompt further emphasizes the key information to help the LLM narrow down its focus and improve the quality of the summary. Beyond the key information, the protocol also instructs the LLM to highlight any extra communication needs with healthcare providers if requested by users. For example, if the patient shared a particular concern when asked "is there anything else that I could help?", the system would present the response as an additional note. This piece usually comes together with each question protocol. **(3) System Setting.** This part instructs the role of the AI assistant (as a text summary tool), its responsibilities (i.e., extracting health information and patient questions), and output format (i.e., concise clinical note structure, Section 3.2.3). The specific wordings are shown in Figure 5. **(4) Conversation Log.** The complete log of the conversation session between the older adult and the VA. **(5) Summary Optimization.** Through our pilot tests, we noticed the need to adopt an in-context learning paradigm to help the LLM generate high-quality summaries. Therefore, this part shows a summary example to ensure the output is well-structured.

4.2.2 Information Highlight Module. The healthcare providers in our interviews also mentioned the importance of having the raw conversation log (Section 3.2.3), enabling them to check the raw data when necessary. We design a highlighting feature to save their effort. The prompt setup is similar to the conversation summary module introduced in Section 4.2.1, with the main difference being the system setting (the responsibility is to find and return the important quotes of the older adult) and output optimization (no longer needed).

4.2.3 Risk Prediction Module. There has been a wide range of research on risk prediction in healthcare (e.g., [42, 63, 86]). We envision our dashboard should also include such a module to help providers allocate limited resources appropriately. For the completeness of our Talk2Care system, we design an LLM-powered risk prediction module. The prompt design is also similar to the conversation summary module in Section 4.2.1, with the main difference being the system setting (the responsibility is to predict the risk level of the older patient: low, moderate, and high) and output optimization (ask for the prediction reasoning). However, we recognize the ethical risks of mis-predictions. Evaluating the prediction performance of the LLM goes beyond the scope of our paper. Additionally, health providers need to implement safety protocols to standardize risk management. Therefore, our evaluation of Talk2Care in Section 5 mainly focuses on the usability of the system design.

4.2.4 Dashboard Interface. We develop our dashboard interface with the React framework. Our design simplifies the interface of existing EHR systems [3, 4] and mainly focuses on the new features we propose. Figure 6 presents the continuous example of the older adult's information, following Figure 3 and Figure 5. Providers can easily navigate to read the conversation summary, the raw conversation logs and highlights, and the communication session history. For each session, they can take follow-up actions or write notes under the summary.

4.2.5 Information Database. The information database of the dashboard is shared with the database of the VA system. It provides the older patient's information, conversation log, and health protocol for prompt input. It also saves providers' action logs or notes for future visualization.

4.2.6 Interaction Flow. Depending on the specific scenarios, the providers first input instructions and protocols into the system, which are used for both the older adult's and the provider's components. When new conversation sessions are stored in the database, the providers are notified. Using the LLM-powered dashboard, they can check the conversation content, take notes, and make decisions on the appropriate follow-up actions.

Combining the two components together, we aim to build Talk2Care as a complete system to facilitate the communication between home-based older adults and healthcare providers.

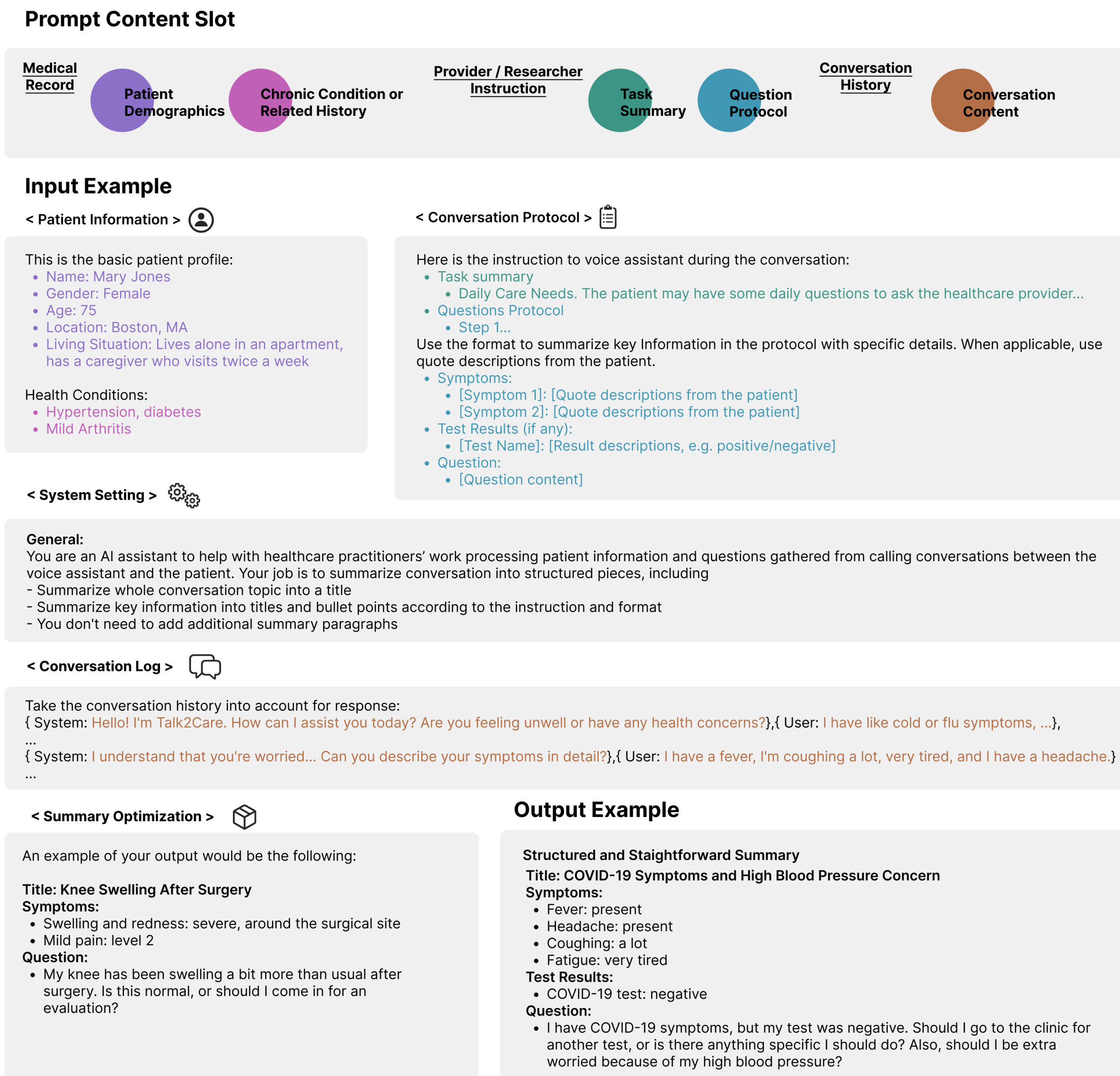

Fig. 5. Prompt Design of Patient-VA Conversation Summary for Healthcare Providers. Similar to Figure 3, the input prompt consists of five parts: 1) patient information, 2) conversation protocol, 3) system setting, 4) conversation log, and 5) summary optimization. For multi-turn conversation, 5) will be repeated for each round of conversation. This figure continues the example of the protocol of daily care.

5 EVALUATION

To evaluate the design of our Talk2Care pilot system for stakeholders, we proposed two specific healthcare scenarios to put Talk2Care in realistic settings (Section 5.1). Then, we conducted two user studies using the two

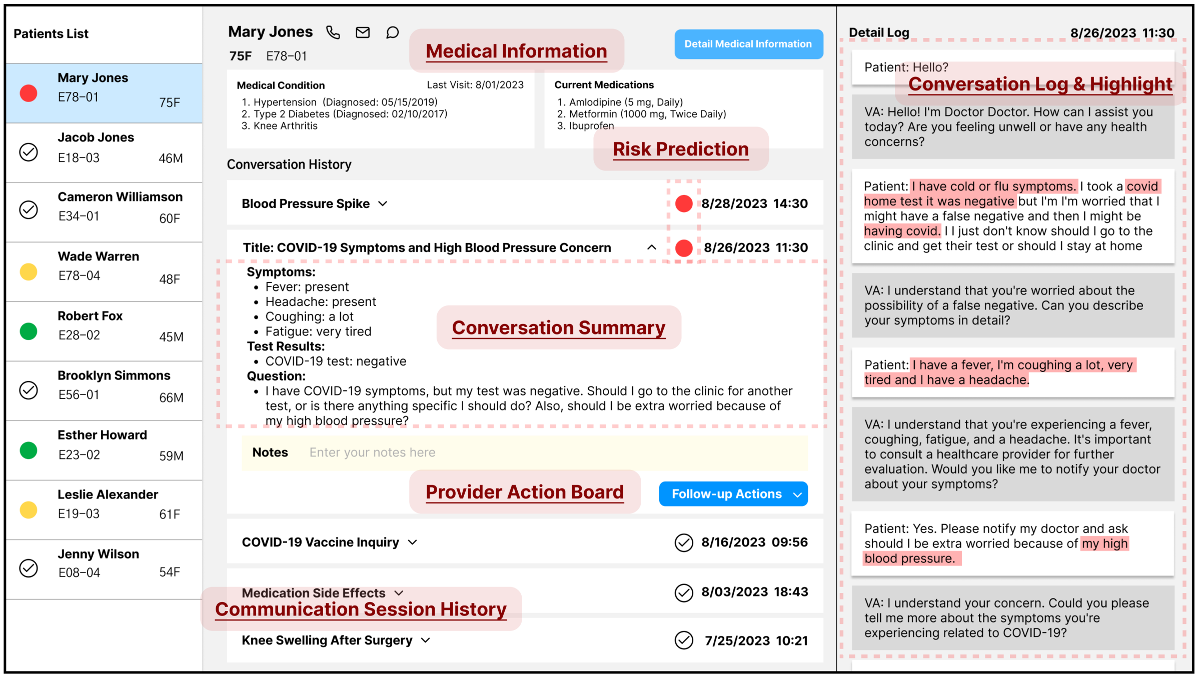

Fig. 6. LLM-powered Dashboard Interface Layout for Healthcare Providers. The dashboard presents the conversation summary (center), raw log, and its highlights (right). The colored dot represents the risk prediction (left and center). Providers can take specific follow-up actions or write down notes under each conversation summary, after which they can mark one session as done.

scenarios with both older adults (Section 5.2) and healthcare providers (Section 5.3) to measure the usability of Talk2Care design.

5.1 Scenario Setup

Based on our interview feedback in Section 3, we designed two scenarios corresponding to the communication needs of older adults at home and healthcare providers, one with older adults as initiators (Section 5.1.1), and the other with providers as initiators (Section 5.1.2).

5.1.1 Scenario A: Daily Care Needs. One of the common communication needs of older adults at home is when they have unexpected health concerns and want to communicate with providers. We designed a character, Mary, a 75-year-old female patient living alone. Mary has been coughing and feeling tired since this morning. She suspects it may be related to COVID-19, but the home-test result turns out to be negative. Mary does not want to travel to the clinic unless necessary, because it takes too much time and energy, and she wants to avoid spreading the potential virus. So she wants to ask for advice on going to urgent care or staying at home. Meanwhile, Mary has high blood pressure, so she is worried about the complex situation and wants to check if there may be any problems.

On the provider side, we designed another character, Tom, a medical office assistant working at a clinic, and his primary responsibility is to answer patient phone calls every day. He gets a number of calls from patients asking about various things, such as lab results, prescriptions, random health-related questions, and requests for appointments. After Tom has all the information from the phone calls, he needs to prioritize them and take the next steps.

5.1.2 Scenario B: Post-Surgery Follow-up. For providers, one common communication need is to follow up with older patients about their symptoms after hospital discharges. Following the character above, John, a 72-year-old male, just went through a small joint surgery on the knee two days ago. He feels good overall, but there is still some pain in the knee. Meanwhile, John has two different kinds of painkillers: aspirin and ibuprofen. He is not sure which one to take.

On the provider side, Emily is a postoperative care nurse working at a busy hospital. Part of her primary responsibility is to follow up with patients after their surgeries, ensuring a smooth recovery process. Emily has to call a number of patients to ask about their current situation and questions. John is one of Emily's patients, so Emily needs to follow up with him about the knee surgery. After Emily has all the information, she will prioritize them and take the next steps.

5.2 User Study 1: Conversational Interface with Older Adults

This user study aims to evaluate the older adult's component of Talk2Care. Our evaluation results indicate that the pilot system can provide good usability (Section 5.2.2), help older adults provide more health information (Section 5.2.2), and offer better mental support (Section 5.2.2).

5.2.1 Methods: Participants and Procedure. With the IRB approval, we recruited the same set of 10 participants from our interview (Table 1). We confirmed that no participants have any hearing or speaking disabilities that may impact the interaction with the VA.

We compared Talk2Care against the most common asynchronous communication method by leaving a message to providers. Specifically, we introduced the two scenarios one by one. After introducing each scenario, older adults were asked to play the role of the character (Mary or John) by writing down (or speaking aloud) the message to leave for providers. Then, they had a conversation with the LLM-powered VA system we implemented, with the same goal of conveying the information to providers. Researchers first gave a brief tutorial on the LLM-powered VA and invited participants to test simple interactions such as invocation. Then the older adults were presented with a sheet to remind them of key invocation phrases and scenario setup. During the interactions with the VA, they were encouraged to engage in natural conversations according to their own language style and add details. After going through the two scenarios, they completed a short evaluation questionnaire and a semi-structured exit interview. Note that the order of the two scenarios was counterbalanced (*e.g.*, participants with odd numbers start with Scenario A and even numbers with Scenario B). Two examples of the conversation log between the older adult and the VA can be found in Appendix B.1 and B.2. Figure 7 shows an example of the user study. The study lasted 25-30 minutes. Participants were compensated $25 for their time.

Our questionnaire includes the System Usability Scale (SUS) [22], the comparison against the basic method of leaving a message, the health support experience, and a few VA-specific experience questions, using a 5-point Likert scale. The SUS is an industrial standard scale to evaluate system usability for users to rate from "strongly disagree" to "strongly agree" on 10 arguments such as whether the system is easy to understand, their need for technical assistance, and learning difficulty [22]. In the interview session, we collected user feedback on what they liked and disliked regarding their interaction experience, as well as their future expectations about the system. The interviews were recorded, transcribed, and processed with Thematic analysis by two researchers. We summarize our findings as follows.

5.2.2 Findings: Good Usability, Description Enhancement, and Mental Support. During the study, all participants were able to successfully complete both scenarios following the instructions. They agreed that the two designed scenarios were close to their daily care needs situations (4.2±1.0 out of 5). Meanwhile, participants found the

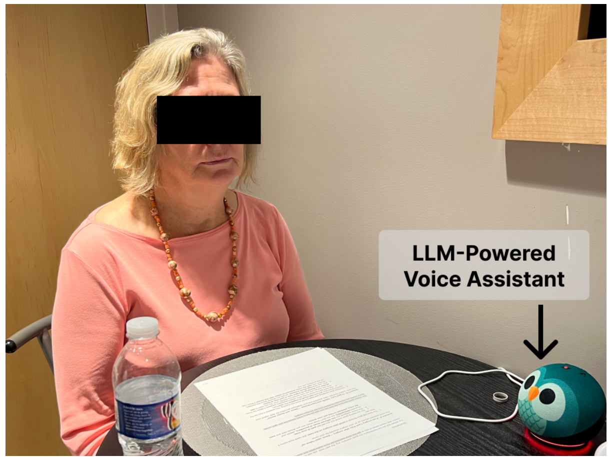

Fig. 7. User Study 3 Scenario with an Older Adult Participant.

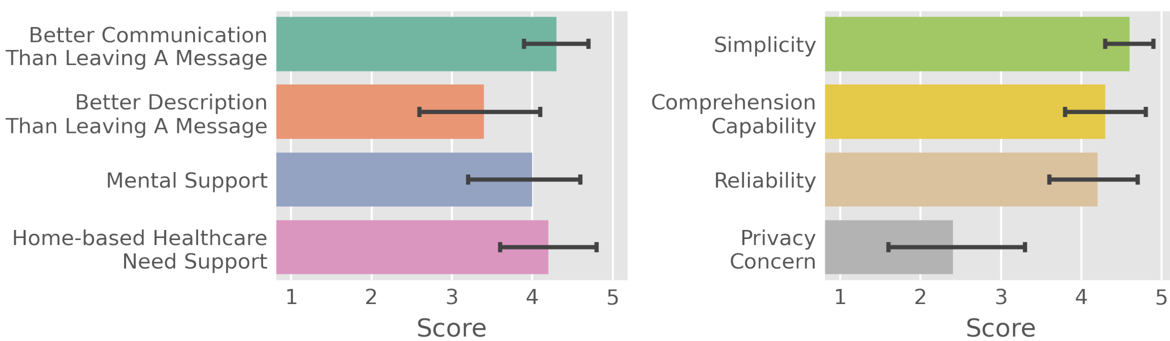

Fig. 8. Subjective Ratings on Different Perspectives of The Usability of The VA System. Error bars show the standard deviation. The average SUS score is 75.5±17.1, indicating good usability.

VA system easy to use. The average SUS score was 75.5±17.1 (out of 100), indicating **good usability.** "*[It's] almost like talking to anybody. I mean, talk to people on the telephone. It's not that different.*" (P10) They also found that the VA system's generated questions were simple to understand (4.6±0.5), and that the system could understand their words accurately (4.3±0.8). Participants agreed that the system was reliable (4.2±1.0) and the overall privacy concern was not high (2.4±1.4). Through our analysis of the conversation logs, we find that the LLM can successfully identify varying ambiguous statements from older adults. For example, in Scenario B, participants described the pain as "*my knees are killing me*" (PP5) or "*I have body aches*" (PP6), and the LLM asked proper follow-up questions about pain level (PP5) or position (PP6).

During the study, participants tried the traditional method of leaving a message, so that they could compare it against the method of talking with the VA. Our results showed that older adults thought the new method could **better support their communication** with providers (4.3±0.7), and **enhance the descriptions of their health information** (3.4±1.3). "*It reminded me of additional symptoms that I went through*" (P9). "*I wouldn't necessarily have thought to include that in my message. [But that is] significant to the doctor.*" (P1) This was a strong advantage of Talk2Care: With the help of natural and interactive conversation, older adults were able to better communicate and give more health information to providers.

As mentioned in Section 3.1.3, older adults thought our system might **provide better mental support**, even though it was intentionally designed to not give specific health advice. This comment was validated after they used our system. Participants gave high ratings on mental support (4.0±1.3) and on the support for their home-based healthcare needs (4.2±1.0). In particular, P2, previously working as a nurse in a local clinic, provided some insights. "*A lot of people are sometimes afraid to ask a doctor. They think it's foolish, so I think that gives them the opportunity to ask without being too threatening by picking up a phone and having to talk to a secretary and go through all of that.*" This provided an interesting effect of AI-mediated communication, which we will further discuss in Section 6.

5.2.3 Findings: Future Expectations and Concerns. Participants had more specific expectations after using the system. A dominating direction was to have the system provide more administrative information, integrate with the healthcare system, and help manage their personal health information. For example, P7 and P9 suggested the VA could provide an estimated time that it takes for the healthcare provider to respond. Participants also mentioned that the system could play the role of "*an entry point of the different specialists*" within the EHR system to minimize the communication effort.

During the study, some participants did experience voice assistant failure such as being interrupted and voice recognition failure. "*I think it's a little awkward, the need to have her complete all her sentences.* " (P6) However, most participants felt confident about adjusting to the interaction mode as long as they have time to practice. "*Once you get into the rhythm, I think it's okay.* " (P6) Yet still, some participants thought the reluctance to adaption for older adults might be a barrier, "because it's a big change". Therefore, a smooth onboarding experience and great usability would be crucial to such an LLM-powered interface for older adults. As mentioned previously, the overall privacy concern was not high (2.4±1.4) among participants. Interestingly, some participants felt that their privacy were better protected interacting with the VA. "*... as opposed to talking to a nurse that could be a next-door neighbor ... there's a little bit more of a barrier between that and humans seeing the information.*" (P9)

5.3 User Study 2: System Interface Study with Healthcare Professionals

Our Study 3 evaluated the first half of the Talk2Care system for older adults. This study focused on evaluating our dashboard design's usability with providers. Our results demonstrate that the pilot system provides good usability (Section 5.3.2), and that it can effectively support providers' communication and information processing (Section 5.3.2).

5.3.1 Methods: Participants and Procedure. Our study was also IRB-approved. We recruited the same set of 9 healthcare providers from our User Study 2 (Table 2). All providers have the experience of using one or multiple EHR systems.

Similar to Study 3, we compared Talk2Care against the most common asynchronous communication method of receiving a message from older patients to providers. We briefly introduced the dashboard interface to providers. After they got familiarized with the system, we introduced the two scenarios to them one by one. Since providers had different responsibilities and specialties, they were not asked to play the role but to suggest the characters (Emily and Tom) about the appropriate actions. In each scenario, they were first asked to give suggestions on the next steps with the traditional communication method, *i.e.*, after receiving patient messages corresponding to the scenario. Then, they were invited to share their screen and navigate through different pages in the dashboard interface by clicking, scrolling, or reviewing information to complete the same goal of processing patient information. In the meantime, they were encouraged to share their thoughts about the interface design. Note that the dashboard designof the two scenarios was the same, except for the different content (scenario A:

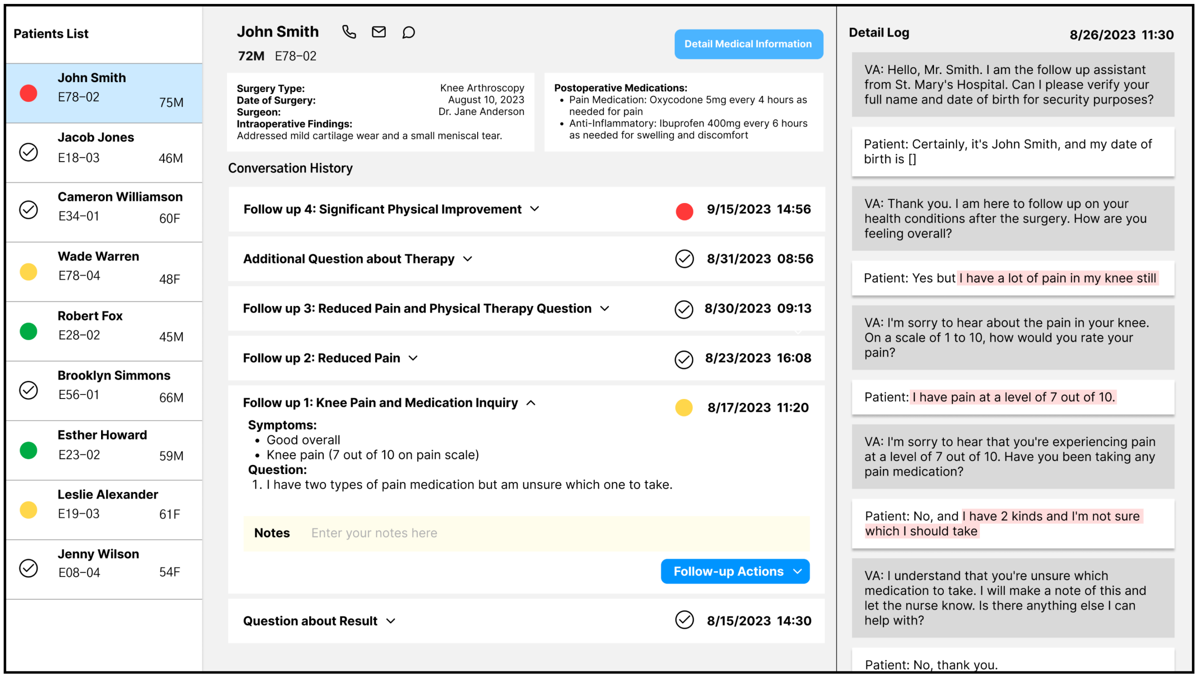

Fig. 9. Dashboard Interface Example for Scenario B: Post-Surgery Follow-up. The example for Scenario A can be found in Figure 6, with the same interface and different scenario-specific content.

Figure 6, scenario B: Figure 9), where we made up patient information for each scenario where the critical cases were as stated in 5.1 After going through the two scenarios (with a counterbalanced order across participants), they completed a short evaluation questionnaire and a semi-structured exit interview. The study lasted 25-30 minutes. Providers were compensated $25 for their time.

Our questionnaire includes the NASA Task Load Index (TLX) [50], the System Usability Scale (SUS) [22], the comparison against the basic method of processing messages, the work support experience, and a few dashboard user experience questions. The TLX questions had a 7-point Likert scale to evaluate aspects of the task load. In questions about demand, effort, or frustration, a lower score indicated a lighter workload and was desirable for the system; in those about performance, a higher score indicated a better outcome. All other questions had a 5-point scale, where higher scores indicated better support. In the interview session, we collected providers' feedback on their experience regarding using the dashboard to facilitate their communication with older adults. We also asked about their future expectations and concerns about the system. Like other studies, two researchers processed the interview transcriptions with Thematic analysis. We summarize our findings below.

5.3.2 Findings: Good Usability, More Information, and AI-Mediated Communication. After navigating through the interface for the two scenarios, providers commented that the system was **easy to understand and convenient to use**. They also agreed that the scenarios and questions were realistic to their actual scenarios, and that the system established a good workflow. The system has an overall SUS score of 85.8±9.8, indicating very good usability. The NASA TLX results were also consistent with their comments (see Figure 10), with low ratings on

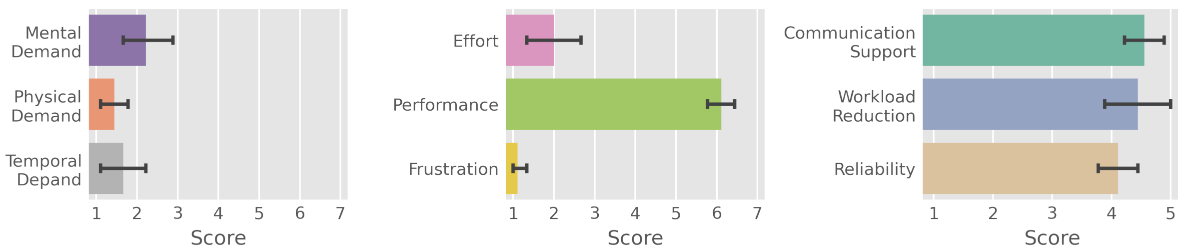

Fig. 10. Subjective Ratings on Different Perspectives of The Usability of The Dashboard System Design. The average SUS score is 85.8±9.8, indicating very good usability.

mental demand (2.2±1.0 out of 7), physical demand (1.4±0.5), temporal demand (1.7±0.9), effort (2.0±1.1), and frustration (1.1±0.3), as well as high ratings on performance (6.1±0.6).

Providers also thought the system was helpful in their communication with older adults (4.6±0.5 out of 5) and offloading their work (4.4±0.9). Six participants mentioned cases where the system could save time for them. For example, PP2, PP3 and PP9 were convinced that the VA system could save their time spent on the phone calling. PP1 also mentioned that the whole workflow might also reduce unnecessary outpatient visits.

Moreover, also consistent with Study 2, more than half of the providers (5 out of 9) mentioned that the LLM-powered dashboard could **assist in triage and help prioritize their tasks**, improving efficiency in the process. Meanwhile, the summary and highlights could help them organize notes for documentation and speed up the review process (PP2, PP3, PP4, PP7, PP9). This was also backed up by the high ratings on the system reliability (4.1±0.6). Providers also favored leveraging the conversation history for their work in long-term, which could be used for clarification, confirmation, and reminder before a patient's visit (PP1, PP3, PP7).

In Study 3, patients commented that talking to the intelligent VA system could help them **reveal more health information** (Section 5.2.2). This was also confirmed by providers in Study 4. Providers agreed that the summary and log contained more information compared to the commonly seen short messages in the traditional patent portal. "*When they're writing, they're probably going to put less detail than when they're talking.*" (PP9) Compared with some questionnaire practices, they also thought that the information could help narrow down the problem to identify the key issue, and get providers more prepared for future visits. With the sufficient information provided, providers also agreed that the communication procedures would be optimized to reduce the back and forth (PP4, PP8).

Interestingly, providers mentioned that the LLM-powered CA could reduce their mental load of talking with patients by playing **the intermediate communication role**. "*[Talking with patients is] pretty soul-sucking, ... and [the nurse] has to be good at keeping their temper and managing people's expectations...*" (PP5)

5.3.3 Findings: Future Expectations and Concerns. In addition to the positive comments, providers also brought up more expectations. Four providers mentioned the need to integrate the system with EHR systems, which was also brought up by older adults in Study 3. Also consistent with older adults' expectations, providers mentioned that the system could offer pre-set valid suggestions and explanations from providers. For instance, examples for rating pain levels should come together with the question querying pain level as pre-set explanations (PP7). Meanwhile, we only covered two protocols for the two scenarios, and providers expected a future intelligent system could incorporate more detailed protocols customized to medical history and key symptoms, and thus collect more detailed health information (PP2, PP3, PP7). For example, PP3 suggested having different protocols for different surgery types and following up with bleeding-related questions, such as recent bleeding time and any effective interventions. With the concerns about AI making mistakes, some providers also suggested that the system should enable them to edit content to correct errors, if any. When "*the human component comes in*" (PP8), they feel they would not *"miss anything"*.

5.4 Finding Summary

In summary, we present the following key findings from our user studies, guiding our discussion in Section 6.

(1) Talk2Care has great accessibility and usability for older adults and their healthcare providers.
(2) The LLM-powered CA could enrich the information communicated and provide mental support for older adults.
(3) The LLM-powered dashboard could promote provider efficiency by offloading and supporting prioritization.
(4) Users identify ethical and privacy concerns as well as desirable features in such LLM-powered systems.

6 DISCUSSION

User Studies 3 and 4 demonstrate the effectiveness of the LLM-powered Talk2Care to facilitate communication between home-based older adults and healthcare providers. In this section, we discuss the effects of AI-mediated communication (Section 6.1), the potential of deployment Talk2Care in real-world scenarios (Section 6.2), and the important ethical concerns and safety considerations of Talk2Care (Section 6.3). We also discuss the limitations of our current work (Section 6.4).

6.1 AI-Empowered Healthcare Work

Existing work has explored how AI may empower healthcare work for providers, focusing on Clinical Decision Support Systems (CDSS) and clinical practices, highlighting how AI may assist human experts throughout medical diagnosis and treatments [41, 73, 76, 115, 127]. Given the limited work in multi-stakeholder contexts [58], our work uncovers AI's capability to address challenges in patient-provider communication. Our interviews and user studies revealed that the LLM-powered system has the strong potential to fill in the gap between older adults' expectations of their healthcare providers and their reality. Firstly, with pre-set protocols and explanations, LLM could satisfy some of the communication needs of patients if their providers are not immediately available, improving responsiveness and supporting triage. Furthermore, LLMs could also address the limitations of existing communication technologies. For example, previous HCI research has also explored strategies and interactive tools for healthcare communication such as embodied agents, voice assistants, and mobile apps, which do not necessarily involve AI and thus follow more fixed interaction flows [13, 31, 32, 99, 106]. Based on these foundations, our study presents the benefits of AI in facilitating communication. With given protocols and interactive conversations, they provide higher adaptability and efficiency for asynchronous communication. Thus, LLM-powered CAs could promote the level of details of asynchronous communication. For healthcare providers, such AI could also lighten their workload in patient-provider communication by taking over basic or repetitive tasks, such as answering phone calls according to a protocol and finding key information in a conversation. Thus, AI might be able to help patients and providers prioritize their communication tasks and maximize communication outcomes within limited availability. Future designers and system builders could utilize AI to work on the tasks in healthcare communication that are time-consuming but could be well structured with instructions.

In multiple stages of our user studies, older adults and providers mentioned how the role of the LLM-powered system can support their communication needs and reduce mental load. Even if they do not receive specific health advice, older adults still feel mentally supported when talking to our VA system (Section 5.2.2), and some might feel less judged compared to when talking to a real healthcare provider (Section 3.2.3). For providers, other than saving their effort and time, such an intelligent system can also reduce their mental load for patient communication (Section 5.3.2). These results suggest the potential of AI as a mediator role in human-human communication, which can help to take over some less favorable parts during the communication and improve both patients' and providers' experience. At the same time, as the AI-powered system promotes the professionalism of providers during communication, such an AI mediator also minimizes harm to patient-provider relationships. This work could serve as an example to support the promising future of AI-mediated healthcare communication [53, 79]. For more personal communication scenarios, such as sharing health information across a family, it remains an open question on how such an AI mediator may affect human relationships.

Despite our limited number of participants, our participant groups have a rather wide range of medical conditions (patients) or areas of expertise (providers), covering different communication needs and challenges. Meanwhile, our patient participants' medical conditions are quite common among the general patient population (*e.g.*, migraine, hypertension). Therefore, we envision that our findings could generalize to patient-provider communication for older adults and potentially other patient populations. We discuss more in future work in Section 6.4.

6.2 Necessary Functions before Real-World Deployment

Our studies suggest that Talk2Care has a strong potential to be deployed in the real world. However, there are a number of features that Talk2Care needs to support and be evaluated before deployment (Section 5.2.3 and 5.3.3).

6.2.1 Integration with EHR Systems. First, both older adults and providers suggested the idea for our system to be integrated with EHR systems, which could enable our system to provide more personalized conversation with older adults with more validated and detailed question protocols, update their health information in the system more seamlessly, support providers with a smoother working flow, and further facilitate the communication between older adults and multiple providers. Admittedly, integration with EHR may come across several obstacles, including demand for budget and resources, handling sensitive information, and interoperability issues [33, 52, 118]. We envision such an integration could potentially start at a specialized clinic and include minimal sensitive data.

6.2.2 Incorporating Long-term Conversation History. Another opportunity to develop a personalized experience is to leverage conversational history in the interaction between patients and LLMs. In our study, since we are using a scenario-based user study setting, we mainly leveraged the conversation history on a per-session scale. Our results have revealed that short-term history (within the same round of conversation) helps the system ask key questions, probe for more health details, and keep the conversation brief. Beyond the scope of the interactions during the study, participants also expressed a strong interest in the possibility of using long-term history (between different conversations) to personalize their experiences. This suggests another direction for our future LLM system. For example, if the user mentions a symptom of pain one day, the LLM CA could follow up in the next few days and inquire if the pain persists. In our current implementation, we have enabled a basic long-term memory mechanism in our system by keeping all conversation records. However, we also recognize the technical difficulties when integrating long-term memory. As pointed out by existing work in LLMs and natural language processing, there are restrictions to the number of tokens input to the LLMs [6], limiting the length of conversation history texts that could be included in LLM prompting [12, 70, 116]. Potential solutions may include using conversation summaries instead of original transcripts for long-term use. Still, future work could further explore the technical solutions to enhance long-term system memory and their effectiveness.

6.2.3 Establishing the Boundary of System Capability. Given the complexity of medical conditions, such systems should have flexibility in handling various health concerns. Some functions of the current system need to be automated, such as selecting the question protocol based on the varying scenarios. The protocols could also be more specific to patients' health conditions or adopt more complicated logic compared to our scenario-based studies. In addition, failure of the CA such as voice recognition issues (Section 5.2.3) [27, 83], should also be addressed to ensure the information accuracy.

However, even with close integration with medical and conversation history, such LLM-powered systems may still come across complicated scenarios that challenge their capabilities in real-world deployment. For example, patients may respond with a mixture of symptoms and questions that do not fit into a pre-set protocol. Patients may also ask about a novel or rare term about which the LLM lacks knowledge. This calls for the LLM's higher flexibility in handling various health concerns, which is an active research topic. Coming across these capability boundaries, the LLM should respond with recognition of the limitations, such as "Sorry, I don't know that one, so I will record your questions and have your provider get back to you.".

6.3 Ethical, Privacy and Safety Concerns of LLM-powered Healthcare Communication

Although Talk2Care shows a number of promising advantages, ethical, privacy and safety concerns must be addressed adequately and thoroughly before any real-world deployment. Our participants mentioned a range of ethical concerns that need to be considered in future work (Section 3.1.4 and 3.2.4).

6.3.1 Ethical Concerns: Ensuring Reliability and Human Involvement. Talk2Care is intentionally designed to not provide health advice due to the concern about AI reliability in healthcare [24, 55, 82]. Nevertheless, it can still make mistakes at multiple stages of communication, such as misunderstanding older adults' meaning of a sentence, misleading older adults when providing their health information, or mis-predicting the severity level so that providers' priorities are biased. Potential solutions include having a separate sanity-check AI agent (thus forming a group of LLM-based agents), requesting human attention when the system is uncertain, or a combination of both.

Similarly, participants also reported reliability concerns regarding LLM's risk predictions, summaries, and highlights. Thus, we emphasize that the human counter-check in our dashboard interface is a crucial step in the proposed workflow, and providers should be informed of the potential risks of LLM results. Meanwhile, patients could also have access to the transcriptions and LLM outputs for double-check. Furthermore, we recommend future work to systematically evaluate the reliability of such predictions and interpretations, thus guiding real-world deployment.

Meanwhile, the impact of such a remote communication system on in-person outpatient visits and patient-provider relationship, as mentioned in participants' concerns, is also unclear. Although providers can ask patients to come in for more tests after the system mediates the communication, this system may reduce older adults' desire to leave home and increase health risks. Similarly, given earlier work on deception and social CAs [100], it is possible that this system could lead to the social isolation of some older adults if they communicate more with the system and less with their family members or providers directly. Future LLM-powered systems should further explore the essential patient-provider interactions to where direct synchronous patient-provider interactions are crucial.

6.3.2 Privacy Concerns: Informed Consent and Interventions. In addition, in future deployment, privacy risk is another important concern of the LLM-powered system, as the system may collect patients' personal information.

As mentioned by participants in interviews (Section 3.1.4 and 5.2.3), future designers and developers should refer to regulations such as HIPPA and existing secured telehealth systems to enable secured access to the database [46, 93], and could potentially utilize natural language processing (NLP) methods to filter critical identifiable information. Meanwhile, in deployment, patients should be fully informed of the data collection of the system. We suggest a system should at least cover the following two steps. Firstly, the system should explicitly explain the data usage and privacy concerns, especially in cases when patients have a lower technical literacy. The system can use either traditional written consent forms, explanations by health providers, or explicit reminders from the system. Secondly, during regular human-CA interactions, the system should regularly remind the patients of risks when sensitive data is collected and suggest practical alternatives.

Much future work is needed to carefully address these concerns to ensure a reliable, safe, and robust healthcare communication system.

6.4 Limitations and Future Work

There are a few limitations in our work. First, our sample sizes for the user studies are limited (10 older adults and 9 healthcare providers). Also, since our participants are recruited from the university participation pool, some of them may have higher education levels, digital literacy, and a better understanding of LLMs than other older adult groups. Therefore, the results obtained from our study may not generalize to other populations. We also acknowledge the limited number and occupation of our provider participants. We selected the provider profiles to focus on clinical communication needs and patient concerns, including other healthcare providers who have more frequent communication and patient management experience may enrich the study insights. However, the provider participants are mainly physicians instead of nurse practitioners or support staff. Thus, we suggest future work include a wide range of both patient and provider populations to further study the generalizability of such systems in real-world deployment.

Second, our current system is a proof-of-concept prototype. The protocols we used in the study need manual effort and can be more detailed. As we discussed in Section 6.2, it needs more future work to achieve a smooth deployment. For example, although our findings suggest that LLMs could enable the system to be customized for each patient, we did not integrate patient information and provider instructions in our user study in a large scale. Resource limitations, privacy, and ethical concerns would arise when scaling up the system. Third, the evaluation of the system mainly focused on system usability and LLM opportunities, but we did not evaluate the performance of different modules in Talk2Care, such as the quality of the generation question or summary, or the accuracy of the risk prediction. Before scaling up the system, there should be a comprehensive and systematic evaluation of the LLM performance in the different modules to ensure reliability and minimize risk to patients.

7 CONCLUSION

In this paper, we designed and implemented Talk2Care, a novel LLM-powered pilot system to facilitate communication between older adults at home and healthcare providers. To better understand the communication challenges and opportunities of LLM-based systems, we first conducted two semi-structured interview studies, one with older adults (N=10), and the other with healthcare providers (N=9). The interview results reveal the need for an AI-based asynchronous communication system and shed light on our design of Talk2Care. Our pilot system consists of two components. The patient component is an LLM-powered VA system that generates high-quality questions to collect health information from older adults and send it to providers for further decision-making. The provider component is a dashboard that summarizes and highlights key information from the patient-VA conversation to save providers' time and assist their analysis process. We then conducted two more user studies to evaluate Talk2Care. Our results showed that Talk2Care may address the challenges mentioned by patients and providers, facilitate their communication, enrich the health information shared by older adults, and reduce the efforts of providers. Our work serves as an example of LLMs' application in the intersection of healthcare and interpersonal communication.

REFERENCES

[1] 2021. Handle Requests Sent by Alexa. https://developer.amazon.com/en-US/docs/alexa/custom-skills/handle-requests-sent-by-alexa.html
[2] 2021. IHMHCOVID-PhoneCallCheckInGuidelines. https://www.pih.org/sites/default/files/lc/MH%20Resource%20Library/PIHMHCOVID-PhoneCallCheckInGuidelines.pdf
[3] 2023. Advanced MD EHR System. https://www.advancedmd.com/
[4] 2023. Epic EHR System. https://www.epic.com/

[5] 2023. Get the Facts on Healthy Aging. https://ncoa.org/article/get-the-facts-on-healthy-aging
[6] 2023. OpenAI Platform. https://platform.openai.com
[7] 2023. Postoperative Instructions for Extractions/Oral Surgery. https://www.aapd.org/globalassets/media/policies_guidelines/r_postsurgery.pdf
[8] 2023. Prompt Engineering Guide. https://www.promptingguide.ai/
[9] Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. 2022. Large Language Models are Few-Shot Clinical Information Extractors. http://arxiv.org/abs/2205.12689 arXiv:2205.12689 [cs].
[10] Lakshmi Arbatti, Abhishek Hosamath, Vikram Ramanarayanan, and Ira Shoulson. 2023. What Do Patients Say About Their Disease Symptoms? Deep Multilabel Text Classification With Human-in-the-Loop Curation for Automatic Labeling of Patient Self Reports of Problems.
[11] Diego Arguello, Ethan Rogers, Grant H. Denmark, James Lena, Troy Goodro, Quinn Anderson-Song, Gregory Cloutier, Charles H. Hillman, Arthur F. Kramer, Carmen Castaneda-Sceppa, and Dinesh John. 2023. Companion: A Pilot Randomized Clinical Trial to Test an Integrated Two-Way Communication and Near-Real-Time Sensing System for Detecting and Modifying Daily Inactivity among Adults >60 Years—Design and Protocol. *Sensors* 23, 4 (Feb. 2023), 2221.
[12] Sanghwan Bae, Donghyun Kwak, Soyoung Kang, Min Young Lee, Sungdong Kim, Yuin Jeong, Hyeri Kim, Sang-Woo Lee, Woomyoung Park, and Nako Sung. 2022. Keep Me Updated! Memory Management in Long-term Conversations. https://doi.org/10.48550/arXiv.2210.08750 arXiv:2210.08750 [cs]
[13] Vince Bartle, Liam Albright, and Nicola Dell. 2023. "This machine is for the aides": Tailoring Voice Assistant Design to Home Health Care Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–19.
[14] Vince Bartle, Janice Lyu, Freesoul El Shabazz-Thompson, Yunmin Oh, Angela Anqi Chen, Yu-Jan Chang, Kenneth Holstein, and Nicola Dell. 2022. "A Second Voice": Investigating Opportunities and Challenges for Interactive Voice Assistants to Support Home Health Aides. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–17.
[15] Andrew Bates, DK Arvind, and Janek Mann. 2011. Wireless monitoring of post-operative respiratory complications. In *Proceedings of the 2nd Conference on Wireless Health*. 1–9.
[16] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.
[17] Caterina Bérubé, Zsolt Ferenc Kovacs, Elgar Fleisch, and Tobias Kowatsch. 2021. Reliability of Commercial Voice Assistants' Responses to Health-Related Questions in Noncommunicable Disease Management: Factorial Experiment Assessing Response Rate and Source of Information. *Journal of Medical Internet Research* 23, 12 (Dec. 2021), e32161.
[18] Johnna Blair and Saeed Abdullah. 2019. Understanding the Needs and Challenges of Using Conversational Agents for Deaf Older Adults. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 161–165. https://doi.org/10.1145/3311957.3359487
[19] Ann Blandford, Janet Wesson, René Amalberti, Raed AlHazme, and Ragad Allwihan. 2020. Opportunities and challenges for telehealth within, and beyond, a pandemic. *The Lancet Global Health* 8, 11 (2020), e1364–e1365.
[20] Cati Boulanger, Adam Boulanger, Lilian de Greef, Andy Kearney, Kiley Sobel, Russell Transue, Z Sweedyk, Paul H Dietz, and Steven Bathiche. 2013. Stroke rehabilitation with a sensing surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1243–1246.
[21] Robin Brewer, Casey Pierce, Pooja Upadhyay, and Leeseul Park. 2022. An Empirical Study of Older Adult's Voice Assistant Use for Health Information Seeking. *ACM Transactions on Interactive Intelligent Systems* 12, 2 (June 2022), 1–32.
[22] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).
[23] Patrick D Brophy. 2017. Overview on the challenges and benefits of using telehealth tools in a pediatric population. *Advances in Chronic Kidney Disease* 24, 1 (2017), 17–21.
[24] Phyllis Butow and Ehsan Hoque. 2020. Using artificial intelligence to analyse and teach communication in healthcare. *The breast* 50 (2020), 49–55.
[25] Julia Cambre, Alex C Williams, Afsaneh Razi, Ian Bicking, Abraham Wallin, Janice Tsai, Chinmay Kulkarni, and Jofish Kaye. 2021. Firefox Voice: An Open and Extensible Voice Assistant Built Upon the Web. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18.
[26] 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 (2023), 33.
[27] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human–Computer Interaction* 37, 8 (May 2021), 729–758. https://doi.org/10.1080/10447318.2020.1841438 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2020.1841438.
[28] Yu Chen, Kingsley Travis Abel, John T Janecek, Yunan Chen, Kai Zheng, and Steven C Cramer. 2019. Home-based technologies for stroke rehabilitation: A systematic review. *International journal of medical informatics* 123 (2019), 11–22.

[29] Amy K Chesser, Nikki Keene Woods, Kyle Smothers, and Nicole Rogers. 2016. Health literacy and older adults: A systematic review. *Gerontology and geriatric medicine* 2 (2016), 2333721416630492.

[30] Chia-Fang Chung. 2017. Supporting patient-provider communication and engagement with personal informatics data. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, Maui Hawaii, 335–338.

[31] 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.

[32] Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A. Munson. 2019. Identifying and Planning for Individualized Change: Patient-Provider Collaboration Using Lightweight Food Diaries in Healthy Eating and Irritable Bowel Syndrome. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 7 (mar 2019), 27 pages. https://doi.org/10.1145/3314394

[33] Maribel Cifuentes, Melinda Davis, Doug Fernald, Rose Gunn, Perry Dickinson, and Deborah J. Cohen. 2015. Electronic Health Record Challenges, Workarounds, and Solutions Observed in Practices Integrating Behavioral Health and Primary Care. *The Journal of the American Board of Family Medicine* 28, Supplement 1 (Sept. 2015), S63–S72. https://doi.org/10.3122/jabfm.2015.S1.150133 Publisher: American Board of Family Medicine Section: Original Research.

[34] George H Crossley, Andrew Boyle, Holly Vitense, Yanping Chang, R Hardwin Mead, and Connect Investigators. 2011. The CONNECT (Clinical Evaluation of Remote Notification to Reduce Time to Clinical Decision) trial: the value of wireless remote monitoring with automatic clinician alerts. *Journal of the American College of Cardiology* 57, 10 (2011), 1181–1189.

[35] Smit Desai and Jessie Chin. 2023. OK Google, Let's Learn: Using Voice User Interfaces for Informal Self-Regulated Learning of Health Topics among Younger and Older Adults. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. https://doi.org/10.1145/3544548.3581507

[36] Xianghua Ding, Yunan Chen, Zhaofei Ding, and Yiwen Xu. 2019. Boundary Negotiation for Patient-Provider Communication via WeChat in China. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 157 (nov 2019), 24 pages. https://doi.org/10.1145/3359259

[37] Karen Donelan, Esteban A Barreto, Sarah Sossong, Carie Michael, Juan J Estrada, Adam B Cohen, Janet Wozniak, and Lee H Schwamm. 2019. Patient and clinician experiences with telehealth for patient follow-up care. *Am J Manag Care* 25, 1 (2019), 40–44.

[38] E Ray Dorsey and Eric J Topol. 2016. State of telehealth. *New England journal of medicine* 375, 2 (2016), 154–161.

[39] Shaun M Eack, Catherine G Greeno, and Bong-Jae Lee. 2006. Limitations of the Patient Health Questionnaire in identifying anxiety and depression in community mental health: many cases are undetected. *Research on social work practice* 16, 6 (2006), 625–631.

[40] Marva V Foster and Kristen A Sethares. 2014. Facilitators and barriers to the adoption of telehealth in older adults: an integrative review. *CIN: Computers, Informatics, Nursing* 32, 11 (2014), 523–533.

[41] David Fraile Navarro, A Baki Kocaballi, Mark Dras, and Shlomo Berkovsky. 2023. Collaboration, not confrontation: Understanding general practitioners' attitudes towards natural language and text automation in clinical practice. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–34.

[42] Michelle Louise Gatt, Maria Cassar, and Sandra C Buttigieg. 2022. A review of literature on risk prediction tools for hospital readmissions in older adults. *Journal of Health Organization and Management* 36, 4 (2022), 521–557.

[43] Erik Grönvall and Nervo Verdezoto. 2013. Beyond self-monitoring: understanding non-functional aspects of home-based healthcare technology. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, Zurich Switzerland, 587–596. https://doi.org/10.1145/2493432.2493495

[44] Meghana Gudala, Mary Ellen Trail Ross, Sunitha Mogalla, Mandi Lyons, Padmavathy Ramaswamy, and Kirk Roberts. 2022. Benefits of, Barriers to, and Needs for an Artificial Intelligence–Powered Medication Information Voice Chatbot for Older Adults: Interview Study With Geriatrics Experts. *JMIR Aging* 5, 2 (April 2022), e32169. https://doi.org/10.2196/32169 Company: JMIR Aging Distributor: JMIR Aging Institution: JMIR Aging Label: JMIR Aging Publisher: JMIR Publications Inc., Toronto, Canada.

[45] Kristen R Haase, Theodore Cosco, Lucy Kervin, Indira Riadi, and Megan E O'Connell. 2021. Older adults' experiences with using technology for socialization during the COVID-19 pandemic: Cross-sectional survey study. *JMIR aging* 4, 2 (2021), e28010.

[46] Joseph L. Hall and Deven McGraw. 2014. For Telehealth To Succeed, Privacy And Security Risks Must Be Identified And Addressed. *Health Affairs* 33, 2 (Feb. 2014), 216–221. https://doi.org/10.1377/hlthaff.2013.0997 Publisher: Health Affairs.

[47] Shane Halloran, Lin Tang, Yu Guan, Jian Qing Shi, and Janet Eyre. 2019. Remote monitoring of stroke patients' rehabilitation using wearable accelerometers. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers*. 72–77.

[48] Danny M. den Hamer, Perry Schoor, Tobias B. Polak, and Daniel Kapitan. 2023. Improving Patient Pre-screening for Clinical Trials: Assisting Physicians with Large Language Models. http://arxiv.org/abs/2304.07396 arXiv:2304.07396 [cs].

[49] Christina N. Harrington, Radhika Garg, Amanda Woodward, and Dimitri Williams. 2022. "It's Kind of Like Code-Switching": Black Older Adults' Experiences with a Voice Assistant for Health Information Seeking. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–15.

[50] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[51] Thomas Hartvigsen, Saadia Gabriel, Hamid Palangi, Maarten Sap, Dipankar Ray, and Ece Kamar. 2022. Toxigen: A large-scale machine-generated dataset for adversarial and implicit hate speech detection. *arXiv preprint arXiv:2203.09509* (2022).

[52] Tsipi Heart, Ofir Ben-Assuli, and Itamar Shabtai. 2017. A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. *Health Policy and Technology* 6, 1 (March 2017), 20–25. https://doi.org/10.1016/j.hlpt.2016.08.002

[53] Jess Hohenstein and Malte Jung. 2018. AI-supported messaging: An investigation of human-human text conversation with AI support. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 1–6.

[54] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–19. https://doi.org/10.1145/3544548.3580688

[55] Oksana Iliashenko, Zilia Bikkulova, and Alissa Dubgorn. 2019. Opportunities and challenges of artificial intelligence in healthcare. In *E3S Web of Conferences*, Vol. 110. EDP Sciences, 02028.

[56] Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q. Vera Liao, Khai Truong, and Shwetak Patel. 2018. FarmChat: A Conversational Agent to Answer Farmer Queries. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 170 (dec 2018), 22 pages. https://doi.org/10.1145/3287048

[57] Badeia Jawhari, Dave Ludwick, Louanne Keenan, David Zakus, and Robert Hayward. 2016. Benefits and challenges of EMR implementations in low resource settings: a state-of-the-art review. *BMC medical informatics and decision making* 16, 1 (2016), 1–12.

[58] 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 *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3581503

[59] Eric L Johnson and Eden Miller. 2022. Remote patient monitoring in diabetes: how to acquire, manage, and use all of the data. *Diabetes Spectrum* 35, 1 (2022), 43–56.

[60] Hala F Kasim, Amina Ibrahim Salih, and Farah Mwafaq Attash. 2023. Usability of telehealth among healthcare providers during COVID-19 pandemic in Nineveh Governorate, Iraq. *Public Health in Practice* 5 (2023), 100368.

[61] Anna Kawakami, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghuidi Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. Improving human-AI partnerships in child welfare: understanding worker practices, challenges, and desires for algorithmic decision support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[62] Jay Kayser, Xu Wang, Zhenke Wu, Asha Dimoji, Xiaoling Xiang, et al. 2023. Layperson-Facilitated Internet-Delivered Cognitive Behavioral Therapy for Homebound Older Adults With Depression: Protocol for a Randomized Controlled Trial. *JMIR Research Protocols* 12, 1 (2023), e44210.

[63] Maura Kennedy, Richard A Enander, Sarah P Tadiri, Richard E Wolfe, Nathan I Shapiro, and Edward R Marcantonio. 2014. Delirium risk prediction, healthcare use and mortality of elderly adults in the emergency department. *Journal of the American Geriatrics Society* 62, 3 (2014), 462–469.

[64] A Baki Kocaballi, Kiran Ijaz, Liliana Laranjo, Juan C Quiroz, Dana Rezazadegan, Huong Ly Tong, Simon Willcock, Shlomo Berkovsky, and Enrico Coiera. 2020. Envisioning an artificial intelligence documentation assistant for future primary care consultations: A co-design study with general practitioners. *Journal of the American Medical Informatics Association* 27, 11 (Nov. 2020), 1695–1704.

[65] Clemens Kruse, Joanna Fohn, Nakia Wilson, Evangelina Nunez Patlan, Stephanie Zipp, Michael Mileski, et al. 2020. Utilization barriers and medical outcomes commensurate with the use of telehealth among older adults: systematic review. *JMIR medical informatics* 8, 8 (2020), e20359.

[66] 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* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 129, 7 pages. https://doi.org/10.1145/3544549.3585614

[67] Tiffany H Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digital health* 2, 2 (2023), e0000198.

[68] Bishal Lamichhane. 2023. Evaluation of chatgpt for nlp-based mental health applications. *arXiv preprint arXiv:2303.15727* (2023).

[69] Elina Laukka, Moona Huhtakangas, Tarja Heponiemi, Sari Kujala, Anu-Marja Kaihlanen, Kia Gluschkoff, and Outi Kanste. 2020. Health care professionals' experiences of patient-professional communication over patient portals: systematic review of qualitative studies. *Journal of Medical Internet Research* 22, 12 (2020), e21623.

[70] Gibbeum Lee, Volker Hartmann, Jongho Park, Dimitris Papailiopoulos, and Kangwook Lee. 2023. Prompted LLMs as Chatbot Modules for Long Open-domain Conversation. In *Findings of the Association for Computational Linguistics: ACL 2023*. 4536–4554. https://doi.org/10.18653/v1/2023.findings-acl.277 arXiv:2305.04533 [cs]

[71] Yuanchun Li, Fanglin Chen, Toby Jia-Jun Li, Yao Guo, Gang Huang, Matthew Fredrikson, Yuvraj Agarwal, and Jason I. Hong. 2017. PrivacyStreams: Enabling Transparency in Personal Data Processing for Mobile Apps. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 76 (sep 2017), 26 pages. https://doi.org/10.1145/3130941

[72] Houjiang Liu, Anubrata Das, Alexander Boltz, Didi Zhou, Daisy Pinaroc, Matthew Lease, and Min Kyung Lee. 2023. Human-centered NLP Fact-checking: Co-Designing with Fact-checkers using Matchmaking for AI. *arXiv preprint arXiv:2308.07213* (2023).

[73] Siru Liu, Aileen P Wright, Barron L Patterson, Jonathan P Wanderer, Robert W Turer, Scott D Nelson, Allison B McCoy, Dean F Sittig, and Adam Wright. 2023. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *Journal of the American Medical Informatics Association* 30, 7 (2023), 1237–1245.

[74] Xi Lu, Yunan Chen, and Daniel A. Epstein. 2021. How Cultural Norms Influence Persuasive Design: A Study on Chinese Food Journaling Apps. In *Designing Interactive Systems Conference 2021*. 619–637.

[75] I Madan and S Williams. 2012. Is pre-employment health screening by questionnaire effective? *Occupational medicine* 62, 2 (2012), 112–116.

[76] Farah Magrabi, Elske Ammenwerth, Jytte Brender McNair, Nicolet F De Keizer, Hannele Hyppönen, Pirkko Nykänen, Michael Rigby, Philip J Scott, Tuulikki Vehko, Zoie Shui-Yee Wong, et al. 2019. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearbook of medical informatics* 28, 01 (2019), 128–134.

[77] Lakmini P Malasinghe, Naeem Ramzan, and Keshav Dahal. 2019. Remote patient monitoring: a comprehensive study. *Journal of Ambient Intelligence and Humanized Computing* 10 (2019), 57–76.

[78] Tami L Mark, Katherine Treiman, Howard Padwa, Kristen Henretty, Janice Tzeng, and Marylou Gilbert. 2022. Addiction treatment and telehealth: review of efficacy and provider insights during the COVID-19 pandemic. *Psychiatric Services* 73, 5 (2022), 484–491.

[79] Hannah Mieczkowski, Jeffrey T Hancock, Mor Naaman, Malte Jung, and Jess Hohenstein. 2021. AI-Mediated Communication: Language Use and Interpersonal Effects in a Referential Communication Task. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 1–14.

[80] Tracy L Mitzner, Jenay M Beer, Sara E McBride, Wendy A Rogers, and Arthur D Fisk. 2009. Older adults' needs for home health care and the potential for human factors interventions. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 53. SAGE Publications Sage CA: Los Angeles, CA, 718–722.

[81] 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*. 205–212.

[82] Jessica Morley, Caio CV Machado, Christopher Burr, Josh Cowls, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. 2020. The ethics of AI in health care: a mapping review. *Social Science & Medicine* 260 (2020), 113172.

[83] Grazia Murtarelli, Anne Gregory, and Stefania Romenti. 2021. A conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots. *Journal of Business Research* 129 (May 2021), 927–935. https://doi.org/10.1016/j.jbusres.2020.09.018

[84] Varun Nair, Elliot Schumacher, and Anitha Kannan. 2023. Generating medically-accurate summaries of patient-provider dialogue: A multi-stage approach using large language models. http://arxiv.org/abs/2305.05982 arXiv:2305.05982 [cs].

[85] 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).

[86] Rónán O'Caoimh, Nicola Cornally, Elizabeth Weathers, Ronan O'Sullivan, Carol Fitzgerald, Francesc Orfila, Roger Clarnette, Constança Paúl, and D William Molloy. 2015. Risk prediction in the community: A systematic review of case-finding instruments that predict adverse healthcare outcomes in community-dwelling older adults. *Maturitas* 82, 1 (2015), 3–21.

[87] OpenAI. 2022. Introducing chatgpt. https://openai.com/blog/chatgpt

[88] Shreya Pandit, San Tran, Yiwen Gu, Elham Saraee, Frederick Jansen, Saurabh Singh, Shirene Cao, Arezoo Sadeghi, Eugenia Shandelman, Terry Ellis, and Margrit Betke. 2019. ExerciseCheck: a scalable platform for remote physical therapy deployed as a hybrid desktop and web application. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, Rhodes Greece, 101–109. https://doi.org/10.1145/3316782.3321537

[89] Sun Young Park, So Young Lee, and Yunan Chen. 2012. The effects of EMR deployment on doctors' work practices: A qualitative study in the emergency department of a teaching hospital. *International journal of medical informatics* 81, 3 (2012), 204–217.

[90] Shachaf Poran, Gil Amsalem, Amit Beka, and Dmitri Goldenberg. 2022. With One Voice: Composing a Travel Voice Assistant from Repurposed Models. In *Companion Proceedings of the Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 383–387. https://doi.org/10.1145/3487553.3524228

[91] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of Intelligent Voice Assistants by Older Adults with Low Technology Use. *ACM Transactions on Computer-Human Interaction* 27, 4 (Aug. 2020), 1–27.

[92] Pallavi Rao and Anirudha Joshi. 2020. Design Opportunities for Supporting Elderly in India in Managing their Health and Fitness Post-COVID-19. In *IndiaHCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction*. ACM, Online India, 34–41. https://doi.org/10.1145/3429290.3429294

[93] Office for Civil Rights (OCR). 2015. HIPAA for Professionals. https://www.hhs.gov/hipaa/for-professionals/index.html Last Modified: 2021-08-16T16:09:01-0400.

[94] Steven I. Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D. Weisz. 2023. The Programmer's Assistant: Conversational Interaction with a Large Language Model for Software Development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 491–514. https://doi.org/10.1145/3581641.3584037

[95] Paul K Rubenstein, Chulayuth Asawaroengchai, Duc Dung Nguyen, Ankur Bapna, Zalán Borsos, Félix de Chaumont Quitry, Peter Chen, Dalia El Badawy, Wei Han, Eugene Kharitonov, et al. 2023. AudioPaLM: A Large Language Model That Can Speak and Listen. *arXiv preprint arXiv:2306.12925* (2023).

[96] Donald L Rubin, John Parmer, Vicki Freimuth, Terry Kaley, and Mumbi Okundaye. 2011. Associations between older adults' spoken interactive health literacy and selected health care and health communication outcomes. *Journal of health communication* 16, sup3 (2011), 191–204.

[97] Hyeyoung Ryu, Soyeon Kim, Dain Kim, Sooan Han, Keeheon Lee, and Younah Kang. 2020. Simple and Steady Interactions Win the Healthy Mentality: Designing a Chatbot Service for the Elderly. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–25. https://doi.org/10.1145/3415223

[98] Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. 2017. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1726–1739.

[99] Woosuk Seo, Ayse G Buyuktur, Sanya Verma, Hyeryoung Kim, Sung Won Choi, Laura Sedig, and Sun Young Park. 2021. Learning from healthcare providers' strategies: Designing technology to support effective child patient-provider communication. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[100] Amanda Sharkey and Noel Sharkey. 2021. We need to talk about deception in social robotics! *Ethics and Information Technology* 23, 3 (Sept. 2021), 309–316. https://doi.org/10.1007/s10676-020-09573-9

[101] Julia Shaver. 2022. The state of telehealth before and after the COVID-19 pandemic. *Primary Care: Clinics in Office Practice* 49, 4 (2022), 517–530.

[102] Hua Shen, Chieh-Yang Huang, Tongshuang Wu, and Ting-Hao'Kenneth' Huang. 2023. ConvXAI: Delivering Heterogeneous AI Explanations via Conversations to Support Human-AI Scientific Writing. *arXiv preprint arXiv:2305.09770* (2023).

[103] James Simpson, Franziska Gaiser, Miroslav Macík, and Timna Breßgott. 2020. Daisy: A Friendly Conversational Agent for Older Adults. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. ACM, Bilbao Spain, 1–3. https://doi.org/10.1145/3405755.3406166

[104] Aaron Smith. 2014. Older adults and technology use. https://www.pewresearch.org/internet/2014/04/03/older-adults-and-technology-use/

[105] Brendan Spillane, Emer Gilmartin, Christian Saam, Benjamin R Cowan, and Vincent Wade. 2018. ADELE: Care and Companionship for Independent Aging.. In *ICAHGCA@ AAMAS*. 18–24.

[106] Vera Stara, Benjamin Vera, Daniel Bolliger, Lorena Rossi, Elisa Felici, Mirko Di Rosa, Michiel De Jong, and Susy Paolini. 2020. Usability and Acceptance of the Embodied Conversational Agent Anne by People with Dementia and their Caregivers: an exploratory study in home environment settings (Preprint). *JMIR mHealth and uHealth* (Nov. 2020). https://doi.org/10.2196/25891

[107] Gudrun Theile, Carsten Kruschinski, Marlene Buck, Christiane A Müller, and Eva Hummers-Pradier. 2011. Home visits-central to primary care, tradition or an obligation? A qualitative study. *BMC family practice* 12 (2011), 1–12.

[108] Milka Trajkova and Aqueasha Martin-Hammond. 2020. "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376760

[109] Reed V. Tuckson, Margo Edmunds, and Michael L. Hodgkins. 2017. Telehealth. *New England Journal of Medicine* 377, 16 (Oct. 2017), 1585–1592. https://doi.org/10.1056/NEJMsr1503323 Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMsr1503323.

[110] Göran Umefjord, Herbert Sandström, Hans Malker, and Göran Petersson. 2008. Medical text-based consultations on the Internet: a 4-year study. *International journal of medical informatics* 77, 2 (2008), 114–121.

[111] Shlomo Vinker, Michael Weinfass, Lior M Kasinetz, Eliezer Kitai, and Igor Kaiserman. 2007. Web-based question-answering service of a family physician—the characteristics of queries in a non-commercial open forum. *Medical informatics and the Internet in medicine* 32, 2 (2007), 123–129.

[112] Suzanne Vugs, Niels Tjeerd Benjamin Scholte, Panos Markopoulos, and Eelko Ronner. 2023. Clinician Attitudes Towards Telemonitoring for Heart Failure Care: Opportunities for Design Research. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–8.

[113] Bryan Wang, Gang Li, and Yang Li. 2023. Enabling Conversational Interaction with Mobile UI Using Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 432, 17 pages. https://doi.org/10.1145/3544548.3580895

[114] Ding Wang, Santosh D. Kale, and Jacki O'Neill. 2020. Please Call the Specialism: Using WeChat to Support Patient Care in China. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13.

[115] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI Doctor" in Rural China: Tensions and Challenges in AI-Powered CDSS Deployment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[116] Weizhi Wang, Li Dong, Hao Cheng, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, and Furu Wei. 2023. Augmenting Language Models with Long-Term Memory. https://doi.org/10.48550/arXiv.2306.07174 arXiv:2306.07174 [cs]

[117] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2023. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data.

[118] Chunhua Weng, Paul Appelbaum, George Hripcsak, Ian Kronish, Linda Busacca, Karina W Davidson, and J Thomas Bigger. 2012. Using EHRs to integrate research with patient care: promises and challenges. *Journal of the American Medical Informatics Association* 19, 5 (Sept. 2012), 684–687. https://doi.org/10.1136/amiajnl-2012-000878

[119] Ziang Xiao, Q Vera Liao, Michelle Zhou, Tyrone Grandison, and Yunyao Li. 2023. Powering an AI Chatbot with Expert Sourcing to Support Credible Health Information Access. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 2–18.

[120] Ziang Xiao, Xingdi Yuan, Q Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer. 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*. 75–78.

[121] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 3 (2020), 1–37.

[122] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K Villalba, Janine M Dutcher, Michael J Tumminia, Tim Althoff, Sheldon Cohen, Kasey G Creswell, J David Creswell, et al. 2019. Leveraging routine behavior and contextually-filtered features for depression detection among college students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–33.

[123] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S Kuehn, Jeremy F Huckins, Margaret E Morris, et al. 2023. GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–34.

[124] Xuhai Xu, Bingshen Yao, Yuanzhe Dong, Saadia Gabriel, Hong Yu, James Hendler, Marzyeh Ghassemi, Anind K. Dey, and Dakuo Wang. 2023. Mental-LLM: Leveraging Large Language Models for Mental Health Prediction via Online Text Data. arXiv:2307.14385 [cs.HC]

[125] Xuhai Xu, Han Zhang, Yasaman Sefidgar, Yiyi Ren, Xin Liu, Woosuk Seo, Jennifer Brown, Kevin Kuehn, Mike Merrill, Paula Nurius, et al. 2022. GLOBEM Dataset: Multi-Year Datasets for Longitudinal Human Behavior Modeling Generalization. *Advances in Neural Information Processing Systems* 35 (2022), 24655–24692.

[126] Xuhai Xu, Tianyuan Zou, Han Xiao, Yanzhang Li, Ruolin Wang, Tianyi Yuan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K Dey. 2022. TypeOut: leveraging just-in-time self-affirmation for smartphone overuse reduction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[127] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–11.

[128] Li Yunxiang, Li Zihan, Zhang Kai, Dan Ruilong, and Zhang You. 2023. Chatdoctor: A medical chat model fine-tuned on llama model using medical domain knowledge. *arXiv preprint arXiv:2303.14070* (2023).

[129] Dong Zhang, Shimin Li, Xin Zhang, Jun Zhan, Pengyu Wang, Yaqian Zhou, and Xipeng Qiu. 2023. SpeechGPT: Empowering Large Language Models with Intrinsic Cross-Modal Conversational Abilities. http://arxiv.org/abs/2305.11000 arXiv:2305.11000 [cs].

[130] Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. 2023. Exploring ai ethics of chatgpt: A diagnostic analysis. *arXiv preprint arXiv:2301.12867* (2023).

[131] Tamara Zubatiy, Kayci L Vickers, Niharika Mathur, and Elizabeth D Mynatt. 2021. Empowering Dyads of Older Adults With Mild Cognitive Impairment And Their Care Partners Using Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. https://doi.org/10.1145/3411764.3445124

A APPENDIX: INTERVIEW SCRIPTS

A.1 Interview Scripts for Older Adults

We begin with a brief introduction of our research and study procedure.

Background

(1) Could you please tell me a bit more about your health conditions, such as any chronic diseases, surgeries or accidents?
(2) Could you please tell me about your home setting, like do you live alone or with family?

Recent Healthcare Visit Experience

(1) Could you please give a recent example of your healthcare experience with a provider, such as visiting a doctor?
  (a) What difficulties and inconveniences did you encounter?

Patient-Provider Communication

(1) In this process, could you please tell me more about what and when you communicated with a healthcare provider at home or seeked their support? [What questions and needs do you have?]
  (a) If the health providers are not available, what's your alternative? Any other solutions you refer to?

Personalization/Conversation

(1) In this process, do you have any personalized questions or concerns that you feel can only be addressed by communicating with a healthcare provider(e.g. but not by a friend or online searching)?

Attitudes of VA for patient-provider communication

(1) Do you know about voice assistants such as Alexa? What about AI? How about ChatGPT? If not, I can make a brief introduction.
(2) In your case, if there is such an AI voice assistant to help you communicate with a healthcare provider by . . . , do you think it will be helpful? [What do you think about it? What do you like/dislike about it? What other features would you like or dislike?]

Closing Question

(1) Do you have any questions for us about the study?

A.2 Interview Scripts for Healthcare Providers

We begin with a brief introduction of our research and study procedure.

Background

(1) Could you please tell me a bit more about your work responsibilities communicating with patients at home, such as monitoring their health conditions or answering patient questions?

Patient-Provider Communication

(1) In this process, could you please tell me more about what difficulties and challenges you encountered communicating with patients, especially older adults?
(2) For your difficulties, did you seek any support from technologies?
  (a) Automated calling systems?
  (b) Pre-designed surveys?

Attitudes of AI for patient-provider communication

(1) Do you know about voice assistants such as Alexa? What about AI? How about ChatGPT? If not, I can make a brief introduction.

(2) In your case, if there is such an AI voice assistant to help you gather information from your patients by [referring to their difficult cases mentioned above]
Or: In your case, if there is such an AI that summarizes the conversation and highlights the key information, Do you think it will be helpful?
  (a) What do you think about it?
  (b) What do you like/dislike about it?
  (c) What other features would you like or dislike?

Closing Question

(1) Do you have any questions for us about the study?

B APPENDIX: CONVERSATION HISTORY

B.1 Post-Surgery Example of Conversation Log Between The Older Adult and Our Voice Assistant

Voice Assistant	Hello! I'm Talk2Care. How are you feeling after your surgery? Any discomfort or concerns?
Patient	I'm feeling good overall, thanks for asking. But I have a little pain.
Voice Assistant	I'm glad to hear that you're feeling good overall. I'm sorry to hear about the pain. On a scale of 1 to 10, how would you rate your pain?
Patient	I'm not sure. Can you give some examples of how to rate my pain?
Voice Assistant	On a scale of 1 to 10, with 1 being no pain and 10 being the worst pain imaginable, how would you rate your current pain level?
Patient	I would probably rate 2. I think it's not too severe.
Voice Assistant	I understand. On a scale of 1 to 10, with 1 being mild and 10 being severe, you rate your discomfort as a 2. Is that correct?
Patient	Yes, that's correct.
Voice Assistant	Thank you for letting me know. Is the pain more towards the mild side or is it more severe?
Patient	I would say it's more mild, and actually I'm not sure about my painkiller. I have 2 different kinds, but I kind of want to ask my doctor which one I should use.
Voice Assistant	I understand that you have two different painkillers. Would you like me to pass this information to the doctor and have them advise you on which one to use?
Patient	Yes, please. And also I think one of them is aspirin and the other one is like a paste that I can put on my skin. Maybe you can ask my doctor about that?
Voice Assistant	Thank you for letting me know about the medications. I'll make sure to pass that information along to your doctor. Is there anything else I can help you with?
Patient	No, thank you.
Voice Assistant	Goodbye!

B.2 Daily Care Example of Conversation Log Between The Older Adult and Our Voice Assistant

Patient	Hello?
Voice Assistant	Hello! I'm Talk2Care. How can I assist you today? Are you feeling unwell or have any health concerns?
Patient	I have like cold or flu symptoms, I took a covid home test it was negative, but I, I'm not, I'm worried that I might have a false negative and then I might be having covid. I, I just don't know should I go to the clinic and get their test or should I stay at home.
Voice Assistant	I understand that you're worried about the possibility of a false negative. Can you describe your symptoms in detail?
Patient	I have a fever, I'm coughing a lot, very tired, and I have a headache.
Voice Assistant	I understand that you're experiencing a fever, coughing, fatigue, and a headache. It's important to consult a healthcare provider for further evaluation. Would you like me to notify your doctor about your symptoms?
Patient	Yes, please notify my doctor and ask should I be extra worried because of my high blood pressure.
Voice Assistant	I understand your concern. It's important to monitor your symptoms closely. Would you like me to inform your healthcare provider about your symptoms?
Patient	Inform my doctor, thank you.
Voice Assistant	I have informed your doctor about your symptoms. They will contact you directly for further guidance. Is there anything else I can assist you with?
Patient	No, thank you.
Voice Assistant	Goodbye!