"It's like having a friend': Exploring the Opportunity of Large Language Models (LLMs) in Assisting Retirees In Their Daily Lives

Jiaan Li*
School of Design
The Hong Kong Polytechnic University
Hong Kong, Hong Kong SAR, Hong Kong
23126094r@connect.polyu.hk

Yulu Hu*
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology(Guangzhou)
Guangzhou, GuangDong, China
yhu635@connect.hkust-gz.edu.cn

Shihan Fu
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology(Guangzhou)
Guangzhou, GuangDong, China
sfu663@connect.hkust-gz.edu.cn

Zhiqing Wu
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology(Guangzhou)
Guangzhou, GuangDong, China
zwu755@connect.hkust-gz.edu.cn

Pan Hui
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology(Guangzhou)
Guangzhou, GuangDong, China
panhui@hkust-gz.edu.cn

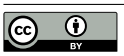
Hailiang Wang
School of Design
The Hong Kong Polytechnic University
Hong Kong, Hong Kong SAR, Hong Kong
hailiang.wang@polyu.edu.hk

Mingming Fan†
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, GuangDong, China
mingmingfan@ust.hk

Abstract

The proportion of older adults is increasing in many countries, and many older adults have reduced work-related and social activities after retirement, which may pose negative challenges to their health. Recent advancements in AI, in particular Large Language Models (LLMs), might have the potential to enrich retirees' daily experiences with their improved natural language understanding and conversation construction abilities. In this work, we sought to explore the potential ways in which retirees might interact with LLMs and their expectations by conducting participatory workshops and a three-week follow-up longitudinal study with 25 retirees to gain insight into their interaction behavior with and perception of LLMs, exemplified by ChatGPT. Our results reveal a range of ways in which older adults feel LLMs may be able to assist them in their daily lives and pinpoint the challenges and concerns faced by this group. Based on the findings, we discussed five key design implications to optimize LLMs-empowered interactions for assisting older adults.

*Both authors contributed equally to this research.
†Corresponding author



CHCHI 2024, Shenzhen, China

ACM ISBN 979-8-4007-1389-7/24/11
https://doi.org/10.1145/3758871.3758911

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**.

Keywords

Older adults, Retiree, Participatory workshop, Follow-up longitudinal study, Human-AI interaction, Large Language Models (LLMs)

ACM Reference Format:
Jiaan Li, Yulu Hu, Shihan Fu, Zhiqing Wu, Pan Hui, Hailiang Wang, and Mingming Fan. 2024. "It's like having a friend': Exploring the Opportunity of Large Language Models (LLMs) in Assisting Retirees In Their Daily Lives. In *Chinese CHI 2024 (CHCHI 2024), November 22–25, 2024, Shenzhen, China.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3758871.3758911

1 Introduction

With the accelerating global trend of aging populations, older adults play an important role in society. Improving their well-being and strengthening social connectedness gradually attract more attention [14, 45]. This population usually experiences physiological, psychological, and social role changes [15, 38]. For example, older adults may face more serious health problems; similarly, many older adults, especially those who have recently retired (i.e., retirees), experience a different lifestyle than they did in their previous lives. Retirement will bring some challenges, such as diminished social networks, changes in daily routines, and a loss of purpose in work [67, 69, 77]. Therefore, it's essential for retirees to re-integrate into society and live a healthy and fulfilling lives.

The development of digital technologies, such as voice assistants [28, 65] and social robots [10] strengthens the connection between older adults and the digital world. These technologies may facilitate simple conversations [58], information search [52], and even task completion when connected to the *Internet of Things* [39]. However, these technologies do not perform well in reasoning and contextualizing continuous dialogues, and the database was limited to specific fields, which restricted their usage among older adults.

The chatbots powered by generative AI (such as OpenAI's ChatGPT, Google's Bard, and Microsoft's Bing Chat) are equipped with Natural Language Generation (NLG) capabilities. Compared to previous intelligent systems, the capabilities of language comprehension [48], contextual awareness [76], and the ability to provide detailed responses [20] have been enhanced significantly. LLMs have been applied to assist older adults across different fields. For example, chronic disease self-management [42], helping individuals with aphasia in word retrieval [34], developing rehabilitation plans [40], and remote diagnostic capabilities [29]. Additionally, LLMs may help older adults in entertainment and psychological health. For instance, Abeer Alessa developed a conversational companion system to mitigate older adults' loneliness and reduce social isolation [3].

The interaction patterns between retirees and large language models (LLMs), such as ChatGPT, and the ways that models may support retirees in their daily activities, as well as the challenges and concerns they face during interaction with LLMs, have not been comprehensively examined. Therefore, we aim to fill in these gaps by answering three research questions (RQs):

- RQ1: How do retirees interact with and perceive LLMs-powered chatbot (e.g., ChatGPT)?
- RQ2: In what aspects might LLMs-powered chatbot (e.g., ChatGPT) assist retirees with their daily lives?
- RQ3: What challenges and concerns do retirees have when using LLMs-powered chatbot (e.g., ChatGPT)?

We conducted a two-part user study, comprising participatory workshops and a follow-up longitudinal study with 25 retirees, to better understand their interaction behavior with and perception of LLM-powered ChatGPT. As intelligent assistants become increasingly ubiquitous, their changing attitude towards ChatGPT indicates that their adaptability to new technology is increasing and is gradually forming an emotional reliance. In addition, we identified four aspects that ChatGPT may provide support for, including cognitive, emotional, creative, and information support. However, we also found challenges when they were interacting with ChatGPT. Based on these findings, we proposed five design implications to provide better services for retired older adults and apply them to the aging population.

2 Related Work

2.1 Aging and Retirement

With the accelerating global trend of aging populations, the roles of older adults in society are becoming increasingly important [21, 36, 59]. Data from the United Nations reports indicate that the global population aged 65 or older is expected to double by 2050. In some developing countries, this age group will account for over one-third of the population [8]. Older adults' roles may transform during retirement [72]. These transitions might lead to psychological fluctuations, with sadness, anxiety, and depression [44]. Additionally, retirement could be seen as an opportunity to explore individual interests and is often viewed as a "second chapter" in life, reducing their workload, mitigating feelings of anxiety and improving their well-being [27, 44, 66, 71]. Therefore, active engagement in retirement activities is helpful for enhancing older adults' healthcare and promoting social engagement, thereby contributing to the promotion of successful aging [62]. With the development of intelligent technologies, the ways of supporting retirees have also changed. Our research aims to address the gap by examining the possibilities of incorporating intelligent technologies into retirees' daily lives and the potential benefits.

2.2 Intelligent Agents for Older Adults

Intelligent agent technology may help older adults manage health challenges. Studies indicate that intelligent agents, such as voice assistants, identify older adults' health needs through communicating with them and provide relevant health advice, such as diabetes management, food, exercise, and medication usage guidelines [5, 6, 11, 53].

As daily living assistants, like Amazon Echo Show or Google Home, intelligent agents help older adults perform daily activities by providing user-friendly services [33, 60]. Instant and personalized responses are helpful for older adults with diminished functional abilities or memory impairment. Additionally, intelligent agents also facilitate older adults' social participation, such as chatbots designed to reduce loneliness and mitigate social isolation [13, 57, 63, 64].

Apart from emotional support, some social robots are used as clinical treatment tools in nursing homes and hospitals [18, 47]. These social robots enhance caregivers' social connection and may help dementia patients mitigate social isolation [43, 52]. Intelligent agents are gradually becoming part of the daily lives of older adults, demonstrating their importance in creating more meaningful lives [49, 50]. Although previous works have shown the intelligent agents' positive influence on older adults, most have been limited to specific contexts. Our work aims to bridge the research gap by examining LLMs' personalization and life-oriented adaptability, and then providing insights into how intelligent systems can be designed to integrate into their daily life.

2.3 Large Language Models (LLMs)

Although intelligent agents have made significant progress, they still have some limitations. Most intelligent agents lack contextual understanding and depend on the data they receive, and only respond to specific commands, which may not meet older adults' expectations [52, 55, 74].

Introducing LLMs such as ChatGPT brings new solutions to address these issues. Different from previous intelligent agents, LLMs are equipped with advanced natural language processing capabilities and can provide more contextual feedback [30]. ChatGPT's training data covers a wide range of fields and delivers in-depth responses, such as cultural, social, economic and others [41]. Moreover, LLMs can progressively infer the user's linguistic habits and optimize responses through continued interaction [16]. This feature

is especially useful for older adults with specific communication needs. Nevertheless, a notable concern is the accuracy of the information. Because natural language processing is not always precise, it may generate persuasive but misleading content that could harm users [24]. Furthermore, some studies have highlighted inherent stereotypes in LLMs; these biases may result in generating inappropriate answers, especially when training data is insufficient [4, 25, 54]. Therefore, it is important to mitigate biases in LLMs and train them to align with ethical standards and societal values, particularly about older adults [31, 61]. Our study aims to compensate for the deficiencies of previous studies by exploring the role of LLMs in potentially empowering retirees and encouraging further research in this emerging area.

While previous research has examined the interaction between humans and LLMs, with limited attention to interactions between older adults, especially retirees and LLMs. Thus, we aim to investigate how older adults interact with and perceive LLMs, how LLMs could assist them in their daily activities, and what challenges they face when using ChatGPT. We will propose design implications for enhancing the design of LLM systems to improve older adults' experiences. Although our study subjects were all retirees, we did not focus on their retirement status or its impact on their interaction behavior with LLMs. We aimed to build a foundation and drive insights by exploring the interaction of a representative sample of older adults with the LLMs.

3 METHOD

We conducted a two-part user study, including (1) participatory workshops and (2) a follow-up longitudinal study, ChatGPT3.5 was used as a tool for this study. Before starting user studies, we organized online conversations with participants to understand the topics they were interested in. This preliminary conversation helped us design collaborative tasks suitable for older adults during the participatory workshops. The research was approved by the university's ethics committee.

3.1 Participants

We recruited participants by posting recruitment advertisements on social media at a local university for the aged. A total of 25 participants signed up for this study, all of whom are retirees. Table 1 presents the participants' demographic information, which includes gender, age, previous work area, understanding level of AI, and usage experience of ChatGPT. They self-assessed their understanding of AI technology on a 5-point Likert scale, with 1 representing no understanding and 5 representing a high level of understanding. The understanding levels were as follows: four participants at Level 1, thirteen at Level 2, and two at Level 3. Only two participants (P12, P16) reported prior experience with GPT.

Table 2 details the information of participants who expressed that they were willing to participate in the follow-up longitudinal study after finishing participatory workshops, comprising nine participants (F=6, M=3) aged between 53 to 62 (μ=57.9, SD= 2.60).

3.2 Procedure

3.2.1 ***Participatory Workshops.*** We conducted seven participatory workshops, each with 3 or 4 participants (in total 25). Each workshop was about one-hour long and consisted of four parts, as shown in Figure 1: introduction, collaborative task, individual task and group interview. We obtained participants' informed consent before the user study, and audio-recorded the study for analysis. Before participatory workshops, we chatted with 15 older adults (11 female, 4 male) through social media platforms, inviting them to select from seven categories - *Health*, *Social*, *Finance*, *Travel*, *Technology*, *Lifestyle* and *Culture* - and rank their interest topics. Results showed that the top three topics were *Health*, *Lifestyle* and *Travel*. Based on these preferences and experiment time limitations, we created two example prompts to present during the introduction. The first prompt, "How to relieve knee pain for 60-year-old adults," relates to *Health*, while the second, "Planning a 7-day self-drive trip to City A," relates the to topic of *Life* and *Traveling*.

Introduction (10 minutes): We first gave a brief introduction of this study. After signing the consent form, participants completed a questionnaire that included demographic information and their self-assessed understanding of ChatGPT. Next, we played a one-minute video to explain what ChatGPT is and several slides to teach how to interact with it. Then, the researcher presented two examples in the following table that older adults are interested in and showed the entire interaction process, including the prompt input and continuous dialogue. We emphasized that ChatGPT can process customized information while also warning about potential risks, such as the possibility of inaccurate or outdated information (i.e., information in ChatGPT up to 2021 when conducting this research). As we observed that participants were unfamiliar with keyboard text input, we introduced the dictation tool to enhance input effectiveness.

Collaborative Task (15 minutes): In this session, all participants in each group would operate one laptop collaboratively. They were required to discuss two provided topics, which were related to their daily life but had some constraints. Two conversation topics are shown in the Table 3. Then, they structured their input prompts and chatted with ChatGPT to gather information, and draw conclusions based on their discussions. Two topics were designated for practice, once participants became familiar with interaction process, they were encouraged to interact freely and ask questions that interested them.

Individual Exploration (30 minutes): Each participant would stay in a different discussion room, provided with a laptop and accompanied by a researcher as an assistant. Participants were asked to chat with ChatGPT on any topic they wished. The researchers would not provide any suggestions except when they encountered difficulties in using ChatGPT.

Group Interview (15 minutes): We conducted group interviews at the end of the study. The interviews were semi-structured and contained questions in three aspects. The first aspect was about their usage experience (e.g., usability and ease of use) and chatting content (e.g., the surprising or unsatisfactory answers). The second aspect was about their attitudes and expectations towards ChatGPT compared with other intelligent chatbots (e.g., Siri) or human beings. The final aspect we collected their opinions on the application scenarios, potential challenges, and functions to be upgraded, etc.

3.2.2 ***Follow-up Longitudinal Study.*** The follow-up longitudinal study lasted one month. Due to regional restrictions, we provide

Table 1: Summary of participants in the user study 1, the offline participatory workshop. Among the 25 participants, 6 were male and 19 were female, aged from 50 to 68. Awareness level of AI is categorised into five levels, from low to high. L1 ——Completely unaware; L2 ——Somewhat aware; L3 ——Moderately aware; L4 ——Very aware; L5 ——Extremely aware.

ID	Sex	Age	Previous Working Area	Awareness level of AI	Usage experience of GPT
P1	M	64	Construction/Engineering	L2	No
P2	F	62	manufacturing industries	L2	No
P3	F	62	manufacturing industries	L2	No
P4	F	55	Finance/Banking	L2	No
P5	F	53	Medical/Well-being	L1	No
P6	F	56	manufacturing industries	L2	No
P7	M	60	manufacturing industries	L2	No
P8	F	62	Government/Public service	L2	No
P9	F	57	manufacturing industries	L1	No
P10	M	65	Construction/Engineering	L2	No
P11	F	62	Construction/Engineering	L1	No
P12	F	62	Government/Public service	L2	Yes
P13	F	68	Finance/Banking	L1	No
P14	F	53	Self-employed	L2	No
P15	F	50	manufacturing industries	L3	No
P16	F	61	manufacturing industries	L2	Yes
P17	M	58	Agriculture/Fisheries	L2	No
P18	F	55	Agriculture/Fisheries	L3	No
P19	F	58	Construction/Engineering	L2	No
P20	F	60	Self-employed	L1	No
P21	F	55	Finance/Banking	L1	No
P22	M	58	Government/Public Service	L1	No
P23	F	57	Construction/Engineering	L1	No
P24	M	57	Government/Public Service	L1	No
P25	F	59	Self-Employed	L1	No

Table 2: Summary of participants in user study 2, the online follow-up longitudinal study.

ID	Sex	Age	Previous Working Area	Awareness level of AI	Usage experience of GPT
P7	M	60	Manufacturing industries	L2	Yes
P11	F	62	Construction/Engineering	L1	Yes
P14	F	53	Self-employed	L2	Yes
P20	F	60	Self-employed	L1	Yes
P21	F	55	Finance/Banking	L1	Yes
P22	M	58	Government/Public Service	L1	Yes
P23	F	57	Construction/Engineering	L1	Yes
P24	M	57	Government/Public Service	L1	Yes
P25	F	59	Self-Employed	L1	Yes

participants with instructions on how to install and use the mobile version of ChatGPT 3.5. To record their interaction patterns, we interviewed them once or twice per week, with each session lasting for 20-30 minutes. Our primary focus of the interviews was to delve into their daily interactions and experiences with ChatGPT 3.5, including their topics of conversation and their opinions of the ChatGPT's answers.

3.3 Data Analysis

In participatory workshops, we collected comprehensive data that included video and audio recordings as well as textual interaction with ChatGPT. The follow-up longitudinal study was conducted remotely, where we collected data consisting of all voice and text information. These data were fully transcribed into text. Two researchers conducted thematic analysis following Braun and Clarke's six-step schema [9]. They thoroughly reviewed all textual content to gain a comprehensive understanding (Step 1). Subsequently, they conductively coded the data (Step 2) to explore participants' behavior interactions, perceptions, expectations and challenges. For example, the codes such as *"Text-to-speech functionality"* and *"Imaging capabilities"* were categorized under the theme *Interaction behavior* (Step 3). Following this, they discussed the coding scheme with the team at regular research meetings and adopted an inductive coding approach to further categorize codes (Step 4). Finally, they reviewed the codes under sub-themes and synthesized them to surround the three topics with initial themes in the research questions. (Step 5 and Step 6).

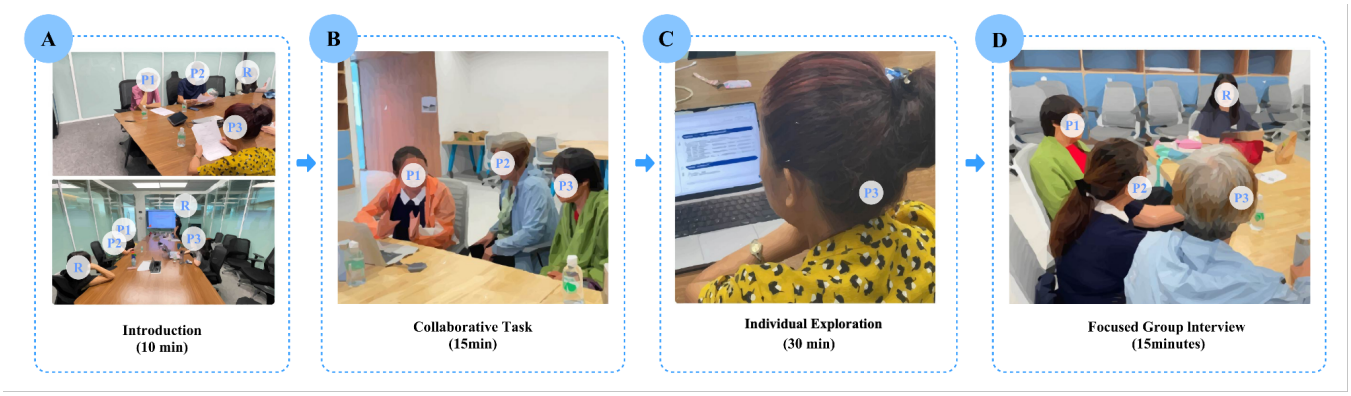


Figure 1: The work flow of the participatory workshops included four parts. A——Introduction, it mainly included filling out demographic information questionnaire and ChatGPT-usage training. B——Collaborative Task, participants shared a computer and collaborated to finish two tasks with ChatGPT. C——Individual Exploration, each participant used a computer for free exploration. D——Group Interview, they shared opinions about the interviewed questions. The circle P means participants, and R means researcher.

Table 3: Tasks for Collaboration

No.	Topics
First Task	You are graduating from a university for older adults, and you would like ChatGPT to help you write a letter to express your gratitude to your dance teacher, preferably in two different versions: one in simplified and one in traditional Chinese.
Second Task	You would like to get outdoors with your family during the three-day festival holiday, so you want ChatGPT to provide you with some suggestions for family-friendly outdoor activities.

4 RESULTS

We first explore participants' interaction behavior and perception in section 4.1 (RQ1) and then the potential ways that LLMs can assist participants from four perspectives in section 4.2 (RQ2). Finally, we present the challenges and concerns associated with using LLMs among participants in section 4.3 (RQ3).

4.1 Interactive Behavior and Perception (RQ1)

We first introduce participants' interactive behavior and perception, including *three patterns that participants prompted with ChatGPT*, *conversational behaviors* and the *perception of interaction.*

4.1.1 ***Prompt patterns with ChatGPT.*** We identified three patterns in which participants prompted ChatGPT: questions, commands, and in-depth dialogue.

Question. Participants proposed clear and straightforward questions that start with inquiry words, such as "How", "What", and "Why". For instance, participants would ask: *"How to make barbecue pork buns?" (P9), "Do people in their sixties need to take calcium supplements?"* (P11) or *"What tips can clean kitchen grease quickly?"* (P6), etc. This prompt pattern allows participants to acquire information in a short time without continuous and repetitive dialogue with ChatGPT.

Command. Commands were generally declarative, direct and imperative, designed to instruct the system to perform the required actions. Participants generally used command prompts to seek detailed instructions, practical plans or a series of tips, etc. For instance, P6 commanded, *"Give me some tips to learn yoga better,"* and P10 commanded, *"Give me some advice on pension investment."* Different from the question-prompt pattern, participants' behavior is goal-oriented within the command-prompt pattern. They expect the system to help them achieve a specific outcome through explicit instruction rather than just acquiring informative answers.

In-Depth Dialogue. This pattern is more complicated and meticulous than the previous two, which typically involve multiple in-depth conversations and feedback loops. Participants usually introduced constraints to the answers by providing a lot of personal information to ChatGPT. A typical example of in-depth dialogue with ChatGPT was about interpersonal communication initiated by P19, she asked : *"How should I communicate with a child who refuses to talk to his parents?"* Following this, she continually provided more details about the child's age, personality and the conflicts that occurred. Through several in-depth dialogues, ChatGPT provided a list of psychotherapy applications that may be helpful to P19.

4.1.2 ***Conversational behavior with ChatGPT.*** Question-and-answering was the predominant pattern of conversation behavior observed between participants and ChatGPT. However, we also identified more diverse behaviors, including *restructuring and refining prompts, customizing questions and dialogue conditioning.*

Restructuring and Refining Prompts. Restructuring refers to participants making multiple modifications and organizing prompts before sending them to ChatGPT. Many participants were aware that prompts strongly influenced the quality of ChatGPT's responses. During the individual exploration phase, participants carefully reviewed and revised the prompts before they submitted them.

In the focus group interview, P19 mentioned that *"Sometimes, I don't know how to express my ideas in my mind."* Some participants would talk to themselves when interacting with ChatGPT, such as *"How can I describe this? Is this OK to say that? "* If answers were not met participants expected, they may feel slightly confused and would revise prompts again in the following conversation.

Customizing Questions and Dialogue Conditions. Participants customized their questions by providing several constraints to find the precise solution step by step. For instance, P16 inputted a prompt, *"Can you help me design three meals a day? I am a 60-year-old man and want to make a one-person recipe based on local tastes for a week"*. Similarly, P13 expected to receive plant advice for establishing a garden, whereby she customized dialogue conditions by adding keywords, such as location, garden area, garden style and plant varieties, to train ChatGPT to generate answers that align with her expectations. Following this, she continued adding more details, such as tree species and winter maturity, to optimize the responses and get more precise recommendations.

4.1.3 ***Perception of Interaction with ChatGPT.*** In this section, we will analyze participants' perceptions of interaction with ChatGPT, consisting of *evolving attitudes*, *humanization* and *comparing ChatGPT with humans and other smart voice assistants*.

The Evolving Attitudes Towards Interaction with ChatGPT. Participants showed curiosity about ChatGPT when they began their attempts and came up with many interesting ideas and questions. Their attitude has gradually changed as they adopt ChatGPT more, mainly attributed to their interaction with and feedback from ChatGPT. Participants would show satisfaction and appreciation when they received the answers they expected. For example, P13 expected ChatGPT to help her draft a speech, she was pleased with the response from ChatGPT:

"The content of this speech is incredibly thorough and logical, saving me a lot of time. Could you send me the draft after the user study?"

Conversely, some participants have become more negative due to unsatisfactory responses. The decreased expectation will further impact their willingness for subsequent participation. In our study, a participant asked ChatGPT questions that he already knew the answers. However, he was skeptical of the credibility of the information when he discovered errors in the answer. For example, P17 asked questions about improving the sharpness of the pictures using Photoshop. However, he noted the errors and said: *"The liquefaction tool in Photoshop is primarily used for shape adjustments, such as modifying the body proportions and movements of a person. It can not improve picture sharpness."*

Furthermore, if participants continuously received unsatisfactory responses, such as incorrect or misleading information, they might lose patience and respond as follows: *"Are you sure?"* or *"You are totally wrong"*. For example, P1 wanted ChatGPT to tell him a fable but received irrelevant responses, which led to his disappointment with ChatGPT, commented: *"Our life experience could help us critically evaluate ChatGPT's answer"*. Therefore, P1 decided to terminate the individual tasks earlier than planned.

Humanizing ChatGPT During Interaction. Participants often treated ChatGPT as human during interactions, following the interpersonal communication norms. On the one side, participants interacted with ChatGPT using polite and formal language, beginning with a formal greeting such as *"Hello"* and using courteous phrases such as *"Please"* and *"Could you"* to start their dialogues. They would express gratitude when the answers met their expectations. On the other hand, if participants were disappointed with responses, they usually showed disappointment to other group members in a polite way rather than directly criticizing ChatGPT, such as *"It seems to have some inaccuracies"*. P13 even attempted to encourage ChatGPT with statements such as: *"Although you can't provide me with a design image draft right now, I'm sure you'll be able to provide me with what I need in the not-too-distant-future, right? Otherwise, I'll be a little disappointed"*.

Comparing ChatGPT with Humans and Other Smart Voice Assistants. During the focus group interviews, participants were encouraged to share their insights on interaction with ChatGPT, humans, and smart voice assistants. Some participants believed that ChatGPT performed better than humans and smart voice assistants in some aspects, but worse in others.

Interacting with humans usually involves social behaviors that may make people feel social pressure. In contrast, interacting with ChatGPT may mitigate pressures. For instance, P15 shared her experiences and stated: *"I can treat ChatGPT as a confidant, sharing things that are not suitable to talk with friends"*. P18 also mentioned that: *"When communicating with ChatGPT, there is no need to worry about its emotions. We can discuss things that are embarrassing to talk about or require social etiquette when communicating with human beings, without having to consider those factors when interacting with it"*. P19 agreed and said: *"If you are communicating with someone, you should notice his/her emotions of the person and respond to their reactions"*. She further highlighted that: *"I will choose ChatGPT to solve my basic life questions if I have a choice"*.

ChatGPT's intelligence has been praised, P13 mentioned that: *"You can only ask a voice assistant one specific question at a time, but you cannot have a consistent and deeper chat. Additionally, it will not give me a summary or conclusion"*. P13 also mentioned that *"For Xiaodu [a voice assistant that retrieves relevant information from search engines], it often says 'this question is too difficult for me to answer', even though the question is not difficult to answer"*.

However, human beings cannot be substituted. ChatGPT can generate text-based responses but it is not capable of deep thinking like the human brain. Therefore, it seldom reflects and poses questions proactively. Such as P17's statement: *"Compared to humans, its responses are too generic. If you were communicating with a person, they would repeatedly ask for details to supplement the conversation"*. He further explained:

> *There's no doubt that humans are better because they can answer more complex questions. For example, if you have electricity problems, you will get better answers from an electrician than consulting ChatGPT. However, an electrician may not be able to provide information about tourist attractions, but ChatGPT can. I think we should inquire about experts' suggestions when it comes to professional knowledge. (P17)*

In addition, ChatGPT can not effectively perform creative writing activities such as writing poetry. P3 made the following prompts to ChatGPT: *"I am a 60-year-old man, I hope in 2024, my son can*

have a baby and succeed in his business. Please give me five couplets." Although ChatGPT answered, it did not follow the traditional principles of poetry, which were superficial and lacked rhythmic symmetry. In addition, when P19 and P3 discussed topics about psychological therapy and medication advice, ChatGPT often hesitated to offer a clear idea, resulting in participants becoming impatient and dropping the conversation.

Moreover, the participants described the responses as *indifferent* (P19), and *lacking warmth* (P22, P24), like P16 said: *"It does not show any emotional change"*. For example, P3 also shared her experience of accidentally waking up the smart voice speaker in her home, she replied *"I did not call you"*, it would reply: *"Alright, my mistake. How about I sing you a song then?"* She believed that the general population would accept a more heartwarming response.

4.2 Potential Ways that LLMs May Assist Older Adults (RQ2)

We identified four potential ways that ChatGPT may assist older adults: 1) Cognitive support, 2) Emotional support, 3) Creative support, and 4) Information support.

4.2.1 ***Cognitive Support.*** We summarized the cognitive support into three parts: learning, comprehension and decision-making.

Learning consists of *self-learning* and *assisting others in learning*. Participants may use ChatGPT to expand their perspectives on history, technology, foreign languages, and other domains. For example, P22 was interested in world political developments and used ChatGPT to enrich his understanding of the history of the Second World War: *"I requested ChatGPT to narrate the history of Second World War; I can engage in a continuous conversation with it and inquire about something I am not familiar with, such as D-Day in the Second World War"*. ChatGPT's contextual awareness is helpful for him because he did not need to enquire questions repeatedly: *"It feels like I'm chatting with someone who can remember our previous questions"*. However, the accuracy and effectiveness of self-learning support can not be guaranteed, as stated by P22, *"But when I asked some specific political questions, it does not have a clear political stance"*.

P20 had self-learned over 30 common English phrases through ChatGPT: *"I input common English phrases from ChatGPT to another English application and follow its pronunciation"*. She also mentioned that: *"I am interested in self-learning English, and I plan to keep up my learning process"*. This statement emphasized that ChatGPT may provide personalized learning support according to individual needs and improve their self-learning ability.

Furthermore, some older adults often find product manuals confusing due to technical jargon and complex language. In a follow-up longitudinal study, P25 shared an experience in which she did not know how to operate a coffee machine: *"I asked about some technical terms, such as 'extraction' and ' compression of coffee grounds'. I think learning some basic coffee-making skills through ChatGPT is very convenient"*. This also shows that ChatGPT may help older adults overcome difficulties and adapt to advanced technologies.

In addition to facilitating self-learning, ChatGPT may also be used as a tool to *assist others in better learning*. P23 shared that her grandson was preparing a speech for the school event, but some wording was inaccurate enough, so she asked ChatGPT to help her grandson revise the drafts. Similarly, P24 mentioned that her Y5 grandson was interested in insects, so she used ChatGPT to help her gain more knowledge about insects and then discuss them with her grandson. These examples highlight the potential of ChatGPT to help participants provide educational and life support for their grandchildren.

Comprehension refers to helping participants to understand complex concepts or knowledge through clear explanations. Older adults often face too-specialized topics that arouse their curiosity and motivate them to seek more information. For example, P15 asked ChatGPT a question related to finance: *"What does it mean that the booking rate for insurance has been reduced from 3.5 to 3?"* ChatGPT provided a professional answer. Similarly, P13 inquired the details of the new technology: *"What is the principle of light wave physiotherapy?"*, P5 was confused with a physiological phenomenon: *"Why do I get dizzy when I see people spinning?"* Ultimately, they gained a deeper understanding of these unfamiliar topics.

Decisions can be made more effectively with the help of ChatGPT, which provides various options and explains the advantages of each choice. P13 sought advice on selecting artificial tears, asking: *"What are some of the best brands of artificial tears?"* P19 inquired about suitable footwear for older adults: *" What kind of shoes are suitable for older people? Recommend me some brands"*. After receiving ChatGPT's recommendations, she found it useful and took a photo of the screen to note those brands. Additionally, P10 and P11 asked for accommodation at a spa hotel in City A and a five-star hotel close to attractions in City B. However, they found these responses not very helpful because they were concerned about the authenticity of the information.

4.2.2 ***Emotional Support.*** ChatGPT can be a companion or audience for some participants in their daily lives. According to the follow-up longitudinal study, some participants are willing to confide in ChatGPT about their worries. Moreover, it may act as an *"emotional confessional"* and provide a private space and help participants open up about their worries.

Emotional Comfort. P20 told us that she quarrelled with her family over personal matters, but she did not want to share this with others:

> *"I disagreed with my daughter-in-law, and we both were a bit angry. I think these things were related to my family and might not be suitable to discuss with others. So I thought that I could confide in ChatGPT about the matter"*.

This indicates that ChatGPT has the potential to create a secure space for emotional conversation particularly for participants who are reluctant to discuss certain matters with others, as they are more willing to share their feelings and thoughts with ChatGPT openly. P20 was inspired by ChatGPT's responses to think about familial relationships and probable solutions rationally. This process reveals a way in which the participants feel that LLMs may help her manage emotions and alleviate negative feelings.

> *"It responded to me in a positive manner and help me analyze the reasons behind the conflicts with my daughter-in-law; I think its support in easing my emotional distress is helpful"*.

Emotional Companion. Although ChatGPT cannot experience emotions like humans, it could convey positive and resonant emotional support. P21 also encountered a similar issue:

> *"Last month, my fish left me, I felt really sad. ChatGPT suggested that I could write a letter to my fish and throw it into the sea. It helps me release some of my sadness, (ChatGPT) like a kind of emotional companion for me."*

Likewise, P13 shared a similar sense, stating: *" ChatGPT can be served as 'emotional confessional' when you can't chat with others or nobody to talk to, you can talk to it".*

4.2.3 ***Creative Support.*** The creative support for older adults mainly focuses on writing, especially content generation (e.g., speech drafts, poems, letters, etc.). P13 was a senior volunteer working in public welfare, and she often gave speeches to promote the use of automated external defibrillators (AEDs). She was delighted with the presentation manuscript drafted by ChatGPT: *"This presentation includes all the highlights of what I want to say!"* Additionally, P13 asked ChatGPT to step into the audience's position to consider: *"Could you help me brainstorm the questions that the audience might potentially inquire about?"* ChatGPT answered that audiences may be interested in learning how to operate an AED and its practical cases. This helped her realise how essential preparing cases is to help audiences understand the importance of AED, and P13 expressed that: *"The presentation and suggestion from ChatGPT is valuable for me and helpful to AED advocacy".*

Furthermore, P25 often posted videos and photos on TikTok to record and share her life, but she found it challenging to write captions for the posts.

> *"Actually, I'm not skilled at expressing myself artistically, and often spend lot of time to contemplate. To my surprise, ChatGPT is like my second brain, because it can understand my meaning."*

This also highlights that ChatGPT may support participants in creative writing, especially in finishing writing tasks with less effort.

4.2.4 ***Information Supports.*** ChatGPT can generate a large amount of information, but older adults mainly focus on topics of healthcare, cooking, fitness, and traveling. This information helps them strengthen their information retrieval skills and make decisions.

Healthcare and Well-Being. Older adults are concerned about health-related topics. They used ChatGPT to understand the causes and prevention of chronic diseases and ask for medication advice. P5 inquired questions related to high blood pressure: *"What causes high differential pressure in hypertension?"* and *"Do large blood pressure differentials cause atherosclerosis?"* Based on ChatGPT's responses, she further explored. P5 described ChatGPT as a physician counselor: *"It will aid me in analyzing my health concerns, I don't need to repeat the issues I encounter".* This indicates that older adults are able to engage in constant and personalized consultations based on their health conditions, which helps them minimize information redundancy.

Cooking. Participants have more time to enhance their cooking skills, explore recipes, and design dietary plans based on their health conditions. In our study, P9 was interested in cooking Chinese wheat-based food. She told us about her experience of making Char Siu Bao (a type of Cantonese bun), explaining that traditional Char Siu Bao is shaped like a blooming flower. She had attempted to use Baidu (a search engine) to learn the skills to make this food, but various search results made it difficult to find an ideal solution. Therefore, she consulted ChatGPT, asking: *"How can I make Char Siu Bao that 'blossoms'?"* ChatGPT provided a clear step-by-step guide and emphasized leaving some space after wrapping the buns and steaming for an extra 20 minutes once the buns have fermented and expanded. P9 commented, *"I hadn't noticed letting the buns ferment before steaming, that may be the reason for my previous failures".*

Fitness. Most participants were interested in learning various types of exercises and used ChatGPT to help them design exercise plans based on their preferences and physical functions. P6 was particularly interested in yoga and hoped to receive exercise guidance through interacting with ChatGPT, *"Could you introduce me to some different types of yoga and corresponding levels of difficulty?"* Interestingly, P6 took photos of responses from ChatGPT and commented, *"Some of the yoga types I haven't even heard of, I'll record them and try them later".* Although these interactions are helpful, they could be improved. For example, P24 inquired about fishing, but he was not satisfied with the responses: *"Text is hard to understand, like actions. I believe video is more suitable".* This feedback indicated that the presentation formats of content generation should be further improved, especially for older adults.

Traveling. Retirees usually have more time for entertainment and personal interests, especially traveling. Some participants provided details about their destination, interests, expectations, and requirements to ChatGPT, enabling it to design personalized travel plans. P9 shared the details about her interests and needs: *"I'd prefer to enjoy natural beauty, could you provide some lesser-known attractions?"* Based on her information, ChatGPT presented customized travel suggestions. However, not all responses fully meet participants' expectations. P14 planned to visit a famous mountain and requested ChatGPT to recommend nearby hotels and provide contact numbers. It was disappointing that part of the information was inaccurate, which frustrated participants: *"I dialed the number, but it is invalid; this information needs to be verified".* This feedback shows the potential of ChatGPT to assist participants in accessing personalized travel information but also indicates the importance of enhancing the timeliness and reliability of information.

4.3 Challenges and Concerns (RQ3)

We identified five key challenges that participants faced during their interaction with ChatGPT. They were categorized into *interaction modality*, *information quality*, *expression style*, *cultural differences and security*.

4.3.1 ***Interaction Modality.*** The current interaction modalities are still difficult for participants. We found that most of the participants preferred using the computer's dictation function to transcribe their speech into text. However, P13 pointed out that the speech-to-text often misunderstood his meaning, which may be attributed to participants' dialects or accents impacting transcription accuracy.

We utilized ChatGPT 3.5 in our study, but it lacked accuracy and the capability to process images. For instance, P16 wanted to

design a blueprint for her new garden, and P1 tried to upload an oracle picture to know its meaning, but ChatGPT could not meet their needs. P16 expressed her thoughts on improvements: *"I hope that in the future, ChatGPT will be able to generate images in various artistic styles and allow us to input images as well".* Although the latest version of ChatGPT allows users to upload images, it may affect the accessibility of older adults. Whether introducing multi-model (images or voice) interactions may unintentionally increase older adults' learning burden requires further exploration of their adaptability to optimized functionalities.

Participants need to put extra effort into adapting to the system's output mode. P9 and P15 preferred to use the text-to-speech function, with P9 stating that *"I do not want to read plenty of text".* P15 stated that: *"In the future, it would be better if ChatGPT could have a customized text-to-speech functionality. It would be extremely helpful for us who have difficulty reading".* Even though the latest ChatGPT includes this feature, it still remains an accessibility issue for older adults.

4.3.2 ***Information Quality.*** The information quality includes three aspects: accuracy, timeliness, and practicality.

Accuracy: ChatGPT occasionally generates incorrect information, even though it seems plausible. For example, P2 asked the question: *"I'm in City A, and I have a three-day mid-autumn vacation. Can you provide me with a travel tip?"* However, ChatGPT mentioned a town that the participant had never heard of, causing P2 to be skeptical and ask: *"Are you sure this town exists?"* After being challenged, ChatGPT revised the answers. P1 also encountered a similar challenge when he tried to search for information about the"E1" error code displayed on the air conditioner screen. He already knew the interpretation of the air conditioner error code before asking questions and expected ChatGPT's responses to match that he already knew. However, instead, it gave inaccurate information. He pointed out: *"The E1 error code does not indicate the indoor temperature sensor failure, your answer is not consistent with reality".*

Timeliness: The corpus of the version of ChatGPT used in this study is up to 2021 and did not collect the most recent data. For example, P5 asked for weather forecasts on a potential hurricane, and P15 requested information about daily stock updates. ChatGPT can not provide the correct answers but still presents outdated details to participants. Another example was when P15 asked about acoustic glass brands, it recommended outdated brands.

Practicality: ChatGPT provided some counter-intuitive answers, so it might not always generate practical solutions to solve the problem. For example, ChatGPT provided cooking guidelines to P15 for stir-fried beef with oyster mushrooms. However, she stated that: *"If I follow its steps to stir-fry the beef, the beef ends up overcooked".* P3 also expressed their view on that: *"It is more like a common cookbook that provides the steps and tips. It can help you finish the cooking process, but it will not tell you how to make it taste better or how to process various ingredients in different situations".*

Additionally, ChatGPT's ability to understand also impacts the practicality. ChatGPT could not cover and satisfy all requirements when participants stated multiple limitations. For instance, P2 asked: *"I am 60 years old now, I need to write a New Year's message to my best friend who is 62. I hope she is healthy in the new year. I hope this message is in line with the Chinese expression style".* While P2 evaluated the response as *"too generic"* and stated that *"the important information given to it is not even mentioned".*

4.3.3 ***Expression Style.*** This refers to how content is delivered through language, such as tone and register. However, most participants were not satisfied with ChatGPT's expression style. P1 stated that: *"I'm looking for psychological help because I usually have a lot of dreams when sleeping, it makes me a little worried".* P1 expected that it could feel her concern and respond with a warmer tone, alleviating her anxiety, but instead, she received a cold answer. She thought it was indifferent and replied: *"I don't think your tone is warm enough, I want you to give feedback now as a good friend who listens, rather than just collecting information to give me an unemotional response".* However, in some scenarios when ChatGPT can not generate insightful responses and only meaningless ones, participants thought that the expression style was disrespectful to users. P15 stated: *"It will not give me the direct answer but always says a load of right but useless crap, I do want precise and concise answers".* P10 also mentioned that *"ChatGPT gives too much explanation of general knowledge, which is unnecessary".*

4.3.4 ***Cultural Differences.*** This refers to variations in customs, beliefs and values resulting from diverse cultural backgrounds. Participants mentioned many aspects related to Chinese culture in their questions; however, due to the limitations of the corpus, ChatGPT may not comprehensively infer various cultural contexts.

In this research, ChatGPT could not infer traditional culture very well when one participant asked it to write the "Qiyanjueju", a poetic form comprising four lines of seven syllables, with rhymes in the first, second and fourth lines. In addition, when participants inquired about other questions winthin particular cultural contexts, such as lanterns in the Yuanmingyuan (P11), and therapeutic medicine (P18), ChatGPT's performance was not satisfying. In addition, ChatGPT could not comprehend beliefs from diverse cultural backgrounds. P15 asked ChatGPT for attractions close to a famous mountain, but the replies of ChatGPT were not aligned with participants' cultural interests, *"I think ChatGPT does not know what we need, the attractions are all about natural scenery. But we older adults preferred to visit the cultural sites, such as temples. There are many famous Taoist temples on this mountain which it doesn't mention at all".*

4.3.5 ***Information Security.*** Participants were concerned about information security when using ChatGPT, believing that important information such as phone numbers, ID numbers and addresses should be strictly protected. P15 strongly expressed her concerns, stating *"Wouldn't it be very dangerous if this system had access to my travel and accommodation details? Who will guarantee my safety?"* Additionally, P15 often became annoyed due to receiving advertisement harassment resulting from personal information leaks, and she expressed her concerns: *"It must not exchange information with other platforms. It is often occurrs that my personal information was leaked after I signed up for an application. I receive many annoying calls trying to sell me houses, insurance, and medication".*

Additionally, national security concerns were also raised, P3 thought that national security might have risks if the data were captured by patriots or used to serve military purposes. However,

not all participants were concerned about these risks. A few participants believed that private information was useless since they were ordinary people, with P17 stating: *"We trust this technology, as long as it works for me, I will continue to use it".* P2 also believed that: *"The risks were more about the national level rather than the individual level."*

5 DISCUSSION

Although our study specifically investigated a subset of the older adult group, the findings indicate the ChatGPT's potential value for older adults and provide insights into their interaction behavior, experiences, and expectations of LLMs. In this section, we first place our study within the existing works and then discuss the five design implications for optimizing LLMs for older adults.

5.1 Reflecting Based on Existing Research

Our study further enriches the existing research on the interaction behavior of older adults with LLMs. The existing studies mainly focus on the relationship between LLMs and the non-aging population [7, 17, 32]. However, these studies did not thoroughly explore older adults' specific needs and preferences, especially the behavior and communication style. Our study found that older adults were unfamiliar with ChatGPT's functions and operations at the beginning; however, they gradually improved their communication efficiency and depth with ChatGPT through continued practice and interaction. This also indicated that older adults can overcome technical issues and adapt more effectively to new technology. As they accumulate experience, they could customize dialogue flexibly, thus optimizing their user experience. However, we also noticed that older adults tend to evaluate ChatGPT's performance based on their personal experiences and feelings, and their attitudes toward ChatGPT change negatively when receiving unsatisfactory responses, thereby reducing their willingness to continue using this technology (Session 4.3.1). These findings further suggest that older adults are gradually becoming emerging technology-driven adopters and actively constructing interaction patterns with technology based on their self-perception and preferences.

Furthermore, the integration of emotional intelligence into LLMs has received growing attention [1, 22, 70]. Our study results further enrich the previous research on the emotional interaction expectations of LLMs from older adults' perspectives. When interacting with ChatGPT, older adults often used polite language and adhered to social norms, which embodied an anthropomorphic psychological mechanism. ChatGPT is not only a tool for obtaining information but also creates a potential space for emotional connection. With the widespread application of this technology, the continued interaction between older adults and LLMs may develop into a long-term relationship featuring emotional dependence. This relationship could serve as a form of social buffer, providing older adults with companionship and alleviating their loneliness in later life.

We echo the finding from [2] emphasizing the critical roles of LLMs within educational contexts. Our research indicates that ChatGPT has strong potential to support autonomous learning among the older population and strengthen their autonomy in life. ChatGPT facilitates continuous learning opportunities, especially by meeting their needs for diverse and personalized study options. Compared with previous research, our research has evaluated how older adults utilize ChatGPT to acquire information and address daily issues, such as health consultations, cooking skills and traveling plans. These usage scenarios not only present ChatGPT's potential capabilities in information provision but also in influencing their decisions and behaviors, offering valuable insights for future research on the long-term effects of LLMs.

5.2 Design Implications

5.2.1 Enhancing the Multi-Modal Interaction of LLMs for Older Adults. Our finding revealed that it is a struggle for older adults to express their needs exactly due to memory issues or repetition, which aligned with previous research [51]. Therefore, it is necessary to integrate functions that can recognize older adults' language mode into the LLMs. The developers should optimize algorithms to identify ambiguous or incomplete vocabulary and expression styles that older adults commonly use. Moreover, we also found that older adults' visual impairments may impact their ability to understand information provided by LLMs. We recommend adding audio and video functions with text displays to help them comprehend the information, given that current LLMs are commonly equipped with auditory and textual modalities [46]. In addition, providing specialized multi-model options for older adults, such as enhancing voice interaction for those with poor vision and optimizing visual interaction for those with hearing impairment, would further improve the user experience for older adults when using LLMs.

5.2.2 Improving LLM's Empathy through Emotional Training that Targeting on Older Adults. Our results indicate that ChatGPT is deficient in emotional expression and empathy (Session 4.1.2), which is consistent with the findings of Ma et al., who pointed out LLM may not be able to perceive users' emotions and moods accurately, resulting in insensitive or inappropriate response to users' emotional needs [37]. In addition, Wake et al. mentioned that various emotional labels and datasets will impact the stability of ChatGPT's emotion identification ability [68]. We found that older adults expect LLMs to recognize their emotional states and provide warm emotional responses when interacting with them (Session 4.3.3). In order to achieve this goal, LLMs should resonate with older adults' emotional states and build "emotional connections" [26]. Therefore, we suggest integrating emotional labels and datasets that are specially designed for older adults into LLMs and utilizing fine-tuning techniques [23]. Thus, large amounts of text and speech data from the older populations should be collected and labeled by sentiment labeling experts to ensure rigour and precision [78]. Through training models, make LLMs recognize older adults' subtle changes in emotions and respond to them emotionally.

5.2.3 Localizing LLMs for Cultural Diversity. Our findings show that LLMs may have potential biases (Session 4.3.4), which is in line with Liu's findings; the author emphasized that different cultures and moral values may show various levels of bias [35]. This study further highlights the necessity of LLMs' optimization for older adults with varying cultural backgrounds and life experiences, ensuring LLMs could respect and acknowledge older adults' cultural identities and social values [31]. LLMs can provide information that

aligns with local culture through training models with data adapted to specific regions and cultures. Therefore, we suggest optimizing datasets and conducting localized data training focused on the older adult group. It is worth noting that the models are expected to provide multiple languages and dialects, which is essential for older adults in non-English-speaking countries.

5.2.4 Enhancing the Data Quality and Credibility of LLMs. Our findings align with the results of Chen et al., who showed that the deceptive misinformation generated by LLMs may pose a greater challenge to humans [12]. In our study, we found that several older adults received false and deceptive information, making it difficult to distinguish between reality and fiction (Session 4.3.2). This problem may result from LLMs' fabrication of facts and information [75]. It is reasonable to believe that more older adults will receive misinformation or a fabricated reality with an increased sample size. Furthermore, because older adults' concerns are often related to their daily lives, they have a greater need to acquire timely and accurate information (Session 4.3.2). Therefore, we highlight the need to minimize the risks of fabricated information to older adults, which involves enhancing LLM's data quality and credibility.

5.2.5 Protecting Older Users' Privacy and Security. Our study extends previous research and reveals the LLM's risk of security and privacy violation [19, 56, 73], indicating that older adults also worry about their privacy and security when using ChatGPT (Session 4.3.5). It is important for older adults, especially those unfamiliar with new technologies and the mechanism, to protect their privacy; otherwise, it may exacerbate the black box effect [72]. Thus, we recommend that researchers explore older adults' awareness of personal privacy and security issues because the lack of understanding may become a barrier to their use of LLMs.

6 Limitation and Future Work

Most participants in our study were selected from the university for older adults, which may unintentionally exclude older adults who lack higher education or are unfamiliar with digital technology, potentially leading to bias in our study results. However, this method helps us to ensure participants are equipped with basic learning and cognitive ability, enabling them to provide detailed and insightful feedback on LLMs. In future work, we aim to expand our sample size and explore the interaction between specific but popular social identities – retirement and LLMs. Researchers should further investigate how older adults with different educational backgrounds and life experiences interact with LLMs and evaluate how these factors influence the utilization of LLMs through quantitative analysis. Meanwhile, our study was limited by the short time of the follow-up longitudinal study. However, we compensated by ensuring participants' frequent interaction with LLMs, which helped us guarantee the quality and depth of data. Future work should investigate how older adults might use LLM tools in long-term deployed research.

7 CONCLUSION

In conclusion, this study explores older adults' interactive behavior and perception of older adults, identifying areas where LLMs may support them through four key aspects: cognition, emotion, creativity, and information. Furthermore, we comprehensively understand older adults' challenges and concerns when they use LLMs. We then concluded with five design implications for optimizing LLMs tailored for the older adult population, including enhancing the Multi-modal interaction, improving LLMs' empathy through emotional training, localizing LLMs for cultural diversity, enhancing the data quality and credibility, and protecting older adult users' privacy and security.

Acknowledgments

This work is partially supported by Guangzhou Higher Education Teaching Quality and Teaching Reform Project (No.2024YBJG070), and the Hong Kong Polytechnic University grant (No.P0050656).

References

[1] Hojjat Abdollahi, Mohammad H Mahoor, Rohola Zandie, Jarid Siewierski, and Sara H Qualls. 2022. Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. *IEEE Transactions on Affective Computing* 14, 3 (2022), 2020–2032.

[2] Ibrahim Adeshola and Adeola Praise Adepoju. 2023. The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments* (2023), 1–14.

[3] Abeer Alessa and Hend Al-Khalifa. 2023. Towards Designing a ChatGPT Conversational Companion for Elderly People. *arXiv preprint arXiv:2304.09866* (2023).

[4] Nicklaus Badyal, Derek Jacoby, and Yvonne Coady. 2023. Intentional Biases in LLM Responses. *arXiv preprint arXiv:2311.07611* (2023).

[5] João Balsa, Isa Félix, Ana Paula Cláudio, Maria Beatriz Carmo, Isabel Costa e Silva, Ana Guerreiro, Maria Guedes, Adriana Henriques, and Mara Pereira Guerreiro. 2020. Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. *Journal of Medical Systems* 44 (2020), 1–12.

[6] João Balsa, Pedro Neves, Isa Félix, Mara Pereira Guerreiro, Pedro Alves, Maria Beatriz Carmo, Diogo Marques, António Dias, Adriana Henriques, and Ana Paula Cláudio. 2019. Intelligent virtual assistant for promoting behaviour change in older people with T2D. In *Progress in Artificial Intelligence: 19th EPIA Conference on Artificial Intelligence, EPIA 2019, Vila Real, Portugal, September 3–6, 2019, Proceedings, Part I 19*. Springer, 372–383.

[7] Yasmine Belghith, Atefeh Mahdavi Goloujeh, Brian Magerko, Duri Long, Tom Mcklin, and Jessica Roberts. 2024. Testing, Socializing, Exploring: Characterizing Middle Schoolers' Approaches to and Conceptions of ChatGPT. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[8] David E Bloom, David Canning, and Alyssa Lubet. 2015. Global population aging: Facts, challenges, solutions & perspectives. *Daedalus* 144, 2 (2015), 80–92.

[9] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[10] Cynthia L Breazeal, Anastasia K Ostrowski, Nikhita Singh, and Hae Won Park. 2019. Designing social robots for older adults. *Natl. Acad. Eng. Bridge* 49 (2019), 22–31.

[11] 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 (TiiS)* 12, 2 (2022), 1–32.

[12] Canyu Chen and Kai Shu. 2023. Can llm-generated misinformation be detected? *arXiv preprint arXiv:2309.13788* (2023).

[13] Hongtu Chen, Sue E Levkoff, Helianthe Kort, Quentin A McCollum, and Marcia G Ory. 2023. Technological innovations to address social isolation and loneliness in older adults. *Frontiers in Public Health* 11 (2023), 1139266.

[14] Irina Catrinel Crăciun. 2023. Addressing Loneliness in Midlife and Older Age: Increasing Social Skills and Connectedness. In *Fostering Development in Midlife and Older Age: A Positive Psychology Perspective*. Springer, 319–340.

[15] Hossein Ebrahimi, Mina Hosseinzadeh, Reihaneh Seifi Saray, Marian Wilson, and Hossein Namdar Areshtanab. 2022. Ability of older adults to recognize cognitive changes and its relationship with mental health: a cross-sectional study. *BMC geriatrics* 22, 1 (2022), 402.

[16] Luciano Floridi and Massimo Chiriatti. 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines* 30 (2020), 681–694.

[17] Jie Gao, Simret Araya Gebreegziabher, Kenny Tsu Wei Choo, Toby Jia-Jun Li, Simon Tangi Perrault, and Thomas W Malone. 2024. A Taxonomy for Human-LLM Interaction Modes: An Initial Exploration. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–11.

[18] Meghana Gudala, Mary Ellen Trail Ross, Sunitha Mogalla, Mandi Lyons, Padmavathy Ramaswamy, Kirk Roberts, et al. 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 (2022), e32169.

[19] Maanak Gupta, CharanKumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. 2023. From chatgpt to threatgpt: Impact of generative AI in cybersecurity and privacy. *IEEE Access* (2023).

[20] Abid Haleem, Mohd Javaid, and Ravi Pratap Singh. 2022. An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil transactions on benchmarks, standards and evaluations* 2, 4 (2022), 100089.

[21] Sarah Harper. 2014. Economic and social implications of aging societies. *Science* 346, 6209 (2014), 587–591.

[22] Shreya Havaldar, Sunny Rai, Bhumika Singhal, Langchen Liu, Sharath Chandra Guntuku, and Lyle Ungar. 2023. Multilingual language models are not multicultural: A case study in emotion. *arXiv preprint arXiv:2307.01370* (2023).

[23] Tianyu He, Guanghui Fu, Yijing Yu, Fan Wang, Jianqiang Li, Qing Zhao, Changwei Song, Hongzhi Qi, Dan Luo, Huijing Zou, et al. 2023. Towards a Psychological Generalist AI: A Survey of Current Applications of Large Language Models and Future Prospects. *arXiv preprint arXiv:2312.04578* (2023).

[24] Dirk Hovy and Shrimai Prabhumoye. 2021. Five sources of bias in natural language processing. *Language and Linguistics Compass* 15, 8 (2021), e12432.

[25] Dong Huang, Qingwen Bu, Jie Zhang, Xiaofei Xie, Junjie Chen, and Heming Cui. 2023. Bias assessment and mitigation in llm-based code generation. *arXiv preprint arXiv:2309.14345* (2023).

[26] Yuanhui Huang, Quan Zhou, and Anne Marie Piper. 2025. Designing Conversational AI for Aging: A Systematic Review of Older Adults' Perceptions and Needs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[27] Markus Jokela, Jane E Ferrie, David Gimeno, Tarani Chandola, Martin J Shipley, Jenny Head, Jussi Vahtera, Hugo Westerlund, Michael G Marmot, and Mika Kivimäki. 2010. From midlife to early old age: health trajectories associated with retirement. *Epidemiology (Cambridge, Mass.)* 21, 3 (2010), 284.

[28] Valerie K Jones, Michael Hanus, Changmin Yan, Marcia Y Shade, Julie Blaskewicz Boron, and Rafael Maschieri Bicudo. 2021. Reducing loneliness among aging adults: the roles of personal voice assistants and anthropomorphic interactions. *Frontiers in public health* 9 (2021), 750736.

[29] Jan Gresil Kahambing. 2023. ChatGPT, public health communication and 'intelligent patient companionship'. *Journal of public health* (2023), fdad028.

[30] Dinesh Kalla and Nathan Smith. 2023. Study and Analysis of Chat GPT and its Impact on Different Fields of Study. *International Journal of Innovative Science and Research Technology* 8, 3 (2023).

[31] Atoosa Kasirzadeh and Iason Gabriel. 2023. In conversation with Artificial Intelligence: aligning language models with human values. *Philosophy & Technology* 36, 2 (2023), 1–24.

[32] Charlotte Kobiella, Yarhy Said Flores López, Franz Waltenberger, Fiona Draxler, and Albrecht Schmidt. 2024. " If the Machine Is As Good As Me, Then What Use Am I?"–How the Use of ChatGPT Changes Young Professionals' Perception of Productivity and Accomplishment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.

[33] Jarosław Kowalski, Anna Jaskulska, Kinga Skorupska, Katarzyna Abramczuk, Cezary Biele, Wiesław Kopeć, and Krzysztof Marasek. 2019. Older adults and voice interaction: A pilot study with google home. In *Extended Abstracts of the 2019 CHI Conference on human factors in computing systems*. 1–6.

[34] Aditya kumar Purohit, Aditya Upadhyaya, and Adrian Holzer. 2023. ChatGPT in Healthcare: Exploring AI Chatbot for Spontaneous Word Retrieval in Aphasia. (2023).

[35] Yang Liu, Yuanshun Yao, Jean-Francois Ton, Xiaoying Zhang, Ruocheng Guo Hao Cheng, Yegor Klochkov, Muhammad Faaiz Taufiq, and Hang Li. 2023. Trustworthy LLMs: a Survey and Guideline for Evaluating Large Language Models' Alignment. *arXiv preprint arXiv:2308.05374* (2023).

[36] Wolfgang Lutz and Samir Kc. 2011. Global human capital: Integrating education and population. *Science* 333, 6042 (2011), 587–592.

[37] Zilin Ma, Yiyang Mei, and Zhaoyuan Su. 2023. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. *arXiv preprint arXiv:2307.15810* (2023).

[38] Mike Martin, Martin Grünendahl, and Peter Martin. 2001. Age differences in stress, social resources, and well-being in middle and older age. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 56, 4 (2001), P214–P222.

[39] Giuseppe Mincolelli, Silvia Imbesi, Gian Andrea Giacobone, and Michele Marchi. 2019. Internet of things and elderly: Quantitative and qualitative benchmarking of smart objects. In *Advances in Design for Inclusion: Proceedings of the AHFE 2018 International Conference on Design for Inclusion, July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9*. Springer, 335–345.

[40] Kartik Mittal and Minakshi Dhar. 2023. Use of ChatGPT by physicians to build rehabilitation plans for the elderly: A mini-review of case studies. *Journal of the Indian Academy of Geriatrics* 19, 2 (2023), 86–93.

[41] Chanchal Molla, Lisa Mani, Mohammad Rakibul Islam Bhuiyan, and Rashed Hossain. 2023. Examining the Potential Usages, Features, and Challenges of Using ChatGPT Technology: A PRISMA-Based Systematic. *Migration Letters* 20, S9 (2023), 927–945.

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

[43] Claudia Müller, Richard Paluch, and ABM Hasanat. 2022. Care: A chatbot for dementia care. *Mensch und Computer 2022-Workshopband* (2022).

[44] John W Osborne. 2012. Psychological effects of the transition to retirement. *Canadian Journal of Counselling and Psychotherapy* 46, 1 (2012), 45–58.

[45] Andreia F Paiva, Cláudia Cunha, Gina Voss, and Alice Delerue Matos. 2023. The interrelationship between social connectedness and social engagement and its relation with cognition: A study using SHARE data. *Ageing & Society* 43, 8 (2023), 1735–1753.

[46] Denise C Park, Gary Lautenschlager, Trey Hedden, Natalie S Davidson, Anderson D Smith, and Pamela K Smith. 2002. Models of visuospatial and verbal memory across the adult life span. *Psychology and aging* 17, 2 (2002), 299.

[47] Antoine Piau, Rachel Crissey, Delphine Brechemier, Laurent Balardy, and Fati Nourhashemi. 2019. A smartphone Chatbot application to optimize monitoring of older patients with cancer. *International journal of medical informatics* 128 (2019), 18–23.

[48] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training. (2018).

[49] Lazlo Ring, Barbara Barry, Kathleen Totzke, and Timothy Bickmore. 2013. Addressing loneliness and isolation in older adults: Proactive affective agents provide better support. In *2013 Humaine Association conference on affective computing and intelligent interaction*. IEEE, 61–66.

[50] Lazlo Ring, Lin Shi, Kathleen Totzke, and Timothy Bickmore. 2015. Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces* 9 (2015), 79–88.

[51] Shlomit Rotenberg, Shelley Sternberg, and Adina Maeir. 2020. Where did I put my glasses? The lived experience of older adults seeking medical help for perceived memory problems. *Disability and Rehabilitation* 42, 25 (2020), 3606–3613.

[52] Nicole Ruggiano, Ellen L Brown, Lisa Roberts, C Victoria Framil Suarez, Yan Luo, Zhichao Hao, and Vagelis Hristidis. 2021. Chatbots to support people with dementia and their caregivers: systematic review of functions and quality. *Journal of medical Internet research* 23, 6 (2021), e25006.

[53] Jamie Sanders and Aqueasha Martin-Hammond. 2019. Exploring autonomy in the design of an intelligent health assistant for older adults. In *Proceedings of the 24th International conference on intelligent user interfaces: companion*. 95–96.

[54] Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? *arXiv preprint arXiv:2303.17548* (2023).

[55] Anna Schlomann, Hans-Werner Wahl, Peter Zentel, Vera Heyl, Leonore Knapp, Christiane Opfermann, Torsten Krämer, and Christian Rietz. 2021. Potential and pitfalls of digital voice assistants in older adults with and without intellectual disabilities: relevance of participatory design elements and ecologically valid field studies. *Frontiers in psychology* 12 (2021), 684012.

[56] Sakib Shahriar and Kadhim Hayawi. 2023. Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations. *arXiv preprint arXiv:2302.13817* (2023).

[57] Candace L Sidner, Timothy Bickmore, Bahador Nooraie, Charles Rich, Lazlo Ring, Mahni Shayganfar, and Laura Vardoulakis. 2018. Creating new technologies for companionable agents to support isolated older adults. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 3 (2018), 1–27.

[58] Kinga Skorupska, Kamil Warpechowski, Radoslaw Nielek, and Wieslaw Kopec. 2020. Conversational Crowdsourcing for Older Adults: a Wikipedia Chatbot Concept.. In *ECSCW*.

[59] James P Smith. 2009. The impact of childhood health on adult labor market outcomes. *The review of economics and statistics* 91, 3 (2009), 478–489.

[60] Ewan Soubutts, Amid Ayobi, Rachel Eardley, Roisin McNaney, Kirsten Cater, and Aisling Ann O'Kane. 2022. Amazon Echo Show as a Multimodal Human-to-Human Care Support Tool within Self-Isolating Older UK Households. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[61] Yidan Tang. 2023. Understanding Societal Values of ChatGPT. (2023).

[62] Barbra Teater and Jill M Chonody. 2020. How do older adults define successful aging? A scoping review. *The International Journal of Aging and Human Development* 91, 4 (2020), 599–625.

[63] Christiana Tsiourti, Emilie Joly, Cindy Wings, Maher Ben Moussa, and Katarzyna Wac. 2014. Virtual assistive companions for older adults: qualitative field study and design implications. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*. 57–64.

[64] Stefano Valtolina and Liliana Hu. 2021. Charlie: A chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness. In *CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter*. 1–5.

[65] Stefano Valtolina and Mattia Marchionna. 2021. Design of a Chatbot to Assist the Elderly. In *International Symposium on End User Development*. Springer, 153–168.
[66] Iris Van der Heide, Rogier M van Rijn, Suzan JW Robroek, Alex Burdorf, and Karin I Proper. 2013. Is retirement good for your health? A systematic review of longitudinal studies. *BMC public health* 13 (2013), 1–11.
[67] Hanna van Solinge, Marleen Damman, and Douglas A Hershey. 2021. Adaptation or exploration? Understanding older workers' plans for post-retirement paid and volunteer work. *Work, Aging and Retirement* 7, 2 (2021), 129–142.
[68] Naoki Wake, Atsushi Kanehira, Kazuhiro Sasabuchi, Jun Takamatsu, and Katsushi Ikeuchi. 2023. Bias in Emotion Recognition with ChatGPT. *arXiv preprint arXiv:2310.11753* (2023).
[69] Mo Wang. 2007. Profiling retirees in the retirement transition and adjustment process: examining the longitudinal change patterns of retirees' psychological well-being. *Journal of applied psychology* 92, 2 (2007), 455.
[70] Xuena Wang, Xueting Li, Zi Yin, Yue Wu, and Jia Liu. 2023. Emotional intelligence of large language models. *Journal of Pacific Rim Psychology* 17 (2023), 18344909231213958.
[71] Hugo Westerlund, Jussi Vahtera, Jane E Ferrie, Archana Singh-Manoux, Jaana Pentti, Maria Melchior, Constanze Leineweber, Markus Jokela, Johannes Siegrist, Marcel Goldberg, et al. 2010. Effect of retirement on major chronic conditions and fatigue: French GAZEL occupational cohort study. *Bmj* 341 (2010).
[72] Thomas Wischmeyer. 2020. Artificial intelligence and transparency: opening the black box. *Regulating artificial intelligence* (2020), 75–101.
[73] Xiaodong Wu, Ran Duan, and Jianbing Ni. 2023. Unveiling Security, Privacy, and Ethical Concerns of ChatGPT. *arXiv preprint arXiv:2307.14192* (2023).
[74] Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. 2014. Hands free-care free: elderly people taking advantage of speech-only interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. 203–206.
[75] Jia-Yu Yao, Kun-Peng Ning, Zhen-Hui Liu, Mu-Nan Ning, and Li Yuan. 2023. Llm lies: Hallucinations are not bugs, but features as adversarial examples. *arXiv preprint arXiv:2310.01469* (2023).
[76] Pengwei Zhan, Zhen Xu, Qian Tan, Jie Song, and Ru Xie. 2024. Unveiling the Lexical Sensitivity of LLMs: Combinatorial Optimization for Prompt Enhancement. *arXiv preprint arXiv:2405.20701* (2024).
[77] Yujie Zhan, Mo Wang, and Junqi Shi. 2015. Retirees' motivational orientations and bridge employment: Testing the moderating role of gender. *Journal of Applied Psychology* 100, 5 (2015), 1319.
[78] Sicheng Zhao, Xiaopeng Hong, Jufeng Yang, Yanyan Zhao, and Guiguang Ding. 2023. Toward Label-Efficient Emotion and Sentiment Analysis This article introduces label-efficient emotion and sentiment analysis from the computational perspective, focusing on state-of-the-art methodologies, promising applications, and potential outlooks. *Proc. IEEE* (2023).