Exploring Older Adults Personality Preferences for LLM-powered Conversational Companions

Ajwa Shahid
Department of Computer Science
University of Virginia
Charlottesville, Virginia, USA
ckf2dq@virginia.edu

Jane Chung
Nell Hodgson Woodruff School of Nursing
Emory University
Atlanta, Georgia, USA
jane.chung@emory.edu

Seongkook Heo
Department of Computer Science
University of Virginia
Charlottesville, Virginia, USA
seongkook@virginia.edu

Abstract

Studies have shown promise in using conversational agents (CAs) to reduce loneliness among older adults. Recent advances in large language models (LLMs) have enhanced these systems by enabling more natural, human-like interactions. However, little is known about how CA personalities influence user experiences, despite personality being a key factor in human conversation. To explore this, we developed a smart speaker-based CA powered by an LLM and conducted a two-phase user study consisting of in-lab sessions and home deployments. We investigated how older adults perceive different CA personalities and how these personalities affect their interaction experiences. Our findings show that participants could distinguish between different personality characteristics and had varying preferences for different personalities during both short-term and long-term interactions.

CCS Concepts

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

Keywords

Conversational AI, LLM, Social Companion, Older Adults, Personality in Conversational Agent

ACM Reference Format:
Ajwa Shahid, Jane Chung, and Seongkook Heo. 2025. Exploring Older Adults Personality Preferences for LLM-powered Conversational Companions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3706599.3719976

1 Introduction

Studies have shown that older adults are generally more susceptible to loneliness than other age groups, due to loss of social connections, declining health and mobility that limit social activities, and reduced interactions with family and friends [13]. A study also showed that loneliness has been linked with poor health and wellness outcomes for older adults [4]. As a result, there is growing interest within the research community in developing interventions to address loneliness in this population. One promising area of research in Human-Computer Interaction focuses on conversational agents (CAs) to alleviate loneliness among older adults. Research has shown that such CAs can reduce feelings of loneliness by providing easily accessible social interaction, especially for older adults who are more likely to have limited opportunities for social engagement [8]. Multiple studies have shown some promise in the potential of companion CAs in reducing loneliness among older adults [2, 20]. As this is an emerging field, most prior studies have focused on exploring the use of various CAs by older adults, including voice assistants like Amazon Alexa [33] and chatbots such as ChatGPT [1]. Limited research has explored how personalizing CAs can enhance user experience, with most existing studies focusing on contextualizing agent responses based on user information like preferences or routines [2]. However, many important aspects of personalization remain underexplored. One such aspect is the CA's personality, as psychological studies show that personality traits play a significant role in shaping communication patterns [25].

Building on these insights, this work aims to explore the effects of different personalities of Large language model (LLM)-powered CAs on the user experience and perceptions of older adults to answer the following research questions:

- **RQ1**: How do older adults perceive different personality traits in LLM-powered voice conversational agents?
- **RQ2**: What are the observed effects of different personality traits in LLM-powered voice conversational agents on older adults' experience with using a conversational agent?

To answer these research questions in a naturalistic setting, we implemented a smart speaker device that includes a mini PC, speaker, and microphone so that users can use it in their homes, similar to the setup used in a previous study on the use of CAs [1]. We also developed a CA program that uses an LLM for processing the conversations. The CA system we developed supports three distinct personality profiles, each characterized by a dominant trait: Extroversion, Agreeableness, or Conscientiousness. To understand how CA's personality traits affect older adults' perceptions and experiences, we conducted a two-phase user study. In the first phase, older adults participated in an in-lab session, engaging in a 10-minute conversation with each personality variant of the agent. The second phase involved a deployment study, where the device was set up in participants' homes for a 12-day period. Throughout the study, we examined participants' perceptions and experiences with each personality type using a multi-method approach, including personality assessments, preference ratings, interviews, daily diary entries, and

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719976

analysis of conversation transcripts between the older adults and the agents. We found that most participants preferred an Agreeable CA in shorter conversations while preferring an Extroverted CA for long-term use. We also found that Agreeableness can sometimes be perceived as sycophantic, while Extroversion may come across as insincerely positive. We discuss how these traits can be balanced to make them more suitable for conversational companions. We then offer insights into potential next steps that can help achieve our goal of designing the ideal conversational companions for older adults.

2 Related Work

This section reviews previous work on the use of conversational agents for older adults (Section 2.1) and the methods used to test and embody personality in LLMs (Section 2.2).

2.1 Conversational Agents and Older Adults

Previous studies on CAs have emphasized their potential to improve health, well-being, and socialization for older adults in their daily lives. Researchers have studied how older adults use CAs like text chatbots and voice assistants to access general and health information, as a tool to set alarms, reminders, or manage medicine [5, 14, 22, 26, 28, 29, 33, 39]. Several studies have demonstrated the effectiveness of companion CAs in alleviating loneliness and feelings of isolation among older adults [2, 7, 9, 18, 20, 30]. Researchers have also focused on understanding the specific needs and desires of older adults in companion conversational agents. In a study on the long-term use of voice assistants, specifically Amazon Alexa, as conversational partners, Upadhyay et al. found that older adults valued the assistant's "non-intrusive yet always available" nature and its ability to engage in a humorous conversational style. However, they were frustrated by the inability of these assistants to retain conversational context which resulted in a lack of follow-up questions by the assistant during conversations [33]. This limitation can be addressed by utilizing LLMs, which can retain information from previous conversations through their ability to handle long context windows [10].

There is an emerging body of work dedicated to evaluating LLM-powered CAs for older adults. Khoo et al. implemented and tested a social robot powered by GPT-3 that engaged in short conversations with older adults to support their well-being. Participants enjoyed interacting with the system, but researchers noted that enhancing the robot's personalization could foster a stronger sense of connection for some users [21]. Alessa et al. evaluated a ChatGPT-based CA as a companion for older adults [2]. To personalize interactions, they used prompt engineering to incorporate details about each participant, such as their daily routines and preferences. While human evaluations rated the system highly for fluency, the researchers identified its level of engagement as an area for improvement. In a participatory design study on conversational companion robots powered by LLMs, Irfan et al. highlighted the recurring theme of personalization. Older adults expected the agents to remember past conversations and personalize their responses based on previous interactions with the user [18].

Prior work on evaluating LLM-powered CAs highlights a need for personalizing CAs for older adults, mainly in tailoring agent responses to contextual details about the user. However, personalization encompasses a wide range of possibilities, many of which remain underexplored in existing research. A recent study that addressed some of these aspects is a qualitative investigation by Rodriguez et al. [30], which identified tone, speed, and the level of detail in CAs' responses as key elements that could be personalized to enhance the user experience for older adults. While these aspects are important, there are additional dimensions of personalization that merit exploration. For instance, psychological studies have demonstrated how personality significantly influences conversational quality and interaction [15, 25]. However, little is known about how the personality traits can be simulated or can affect the experiences of older adults using their companion CAs. To address this gap, our study takes the first step in exploring the effects of personalizing the personality traits of CAs by examining how different induced personalities in LLM-powered companion CAs are perceived and influence the user experience of older adults.

2.2 Personality in LLMs

Researchers have demonstrated various methods to successfully induce and evaluate personality traits in LLMs. For instance, Safdari et al. demonstrated that LLMs can be reliably assessed for personality traits and shaped to exhibit human-like personalities through prompt design. This approach used keywords from Goldberg's validated list of adjectives [16], which represent the Big Five [27] personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, also known as the OCEAN model. Jiang et al. introduced the Machine Personality Inventory (MPI), based on the Big Five personality factors, to study LLM behaviors, along with the Personality Prompting (P^2) technique [19]. This prompt-based method uses trait-descriptive keywords from psychological studies to refine naive human-generated prompts into keyword-based prompts. The LLM then generated descriptions of individuals exhibiting the target traits to create tailored personality prompts. This method successfully induced specific personality traits in LLMs, validated through the MPI and human-evaluated vignette tests.

Li et al. presented a framework for studying personality in LLMs, along with a comprehensive psychometric benchmark. Their personality induction tests included P^2 [19], which successfully enhanced the targeted personality traits, as confirmed through personality assessments [24]. Based on successes in prior studies, we decided to use the P^2 method [19] for our study to induce personalities in LLM-powered conversational agents.

3 System Implementation

To facilitate natural interaction patterns during our study, we developed a device that emulates a commonly used device - a conventional smart speaker, but enhanced with an LLM. The device is built using a small PC equipped with an Intel N97 processor and 12GB of RAM, a speaker-microphone device, five control buttons, and an LTE modem, all enclosed in a 3D-printed case (dimensions: W:118 x D:118 x H:84 mm, see Figure 1). The buttons were connected to the PC using Arduino Pro Micro. We developed the conversational agent in Python using OpenAI's GPT4o API as the LLM, OpenAI's TTS-1 as the text-to-speech (TTS), and Whisper as the automatic

Figure 1: Smart speaker device developed for the study.

speaker recognition (ASR) model. To simulate companion-like behavior, we incorporated contextual memory retention across sessions. We achieved this by using GPT-4o to generate structured summaries of each conversation, which were then provided as input context for subsequent sessions as illustrated in 2, enabling the agent to recall and build on previous interactions.

For simulating different personalities, we used the personality prompting method developed by Jiang et al. [19]. Our system supports three distinct personalities, guided by the Big Five personality framework [27]. We chose Agreeableness, Conscientiousness, and Extroversion to create three distinct personality types in conversational agents to study their impact on user engagement and conversational quality, while excluding Neuroticism and Openness. Neuroticism was omitted because it is associated with negative affectivity and is undesirable in a social companion [17]. Openness was excluded because it has a strong positive correlation with Extroversion [3]. Prior research has shown that participants often struggle to differentiate between these two traits in CAs, as their overlapping characteristics can make it difficult to tell them apart [32].

4 Methods

4.1 Phase 1 - Lab Study

The goal of the first phase study is to explore how older adults perceive the personalities of LLM-powered conversational agents and whether they can distinguish among the three designed personalities. Five participants were recruited through posters displayed on community boards at elderly care homes in the local area. Participant demographic details are provided in Table 1. Participants completed a 5-item pre-questionnaire to assess their familiarity with AI-powered conversational agents. Then, the researchers demonstrate how to use the device.

Each participant interacted with CAs embodying three distinct personalities (Extroversion, Agreeableness, and Conscientiousness). To control for order effects, the sequence in which participants interacted with each personality was counter-balanced using the Latin Square method. Building on previous research highlighting storytelling as a particularly engaging topic for older adults in their interactions with social companions [34], participants were asked to recall and share a personal story with the agent. To maintain consistency across interactions with different agents, we asked the participants to tell the same story. Interactions lasted 10 minutes per agent, and participants were instructed to keep the story content consistent across the sessions. After each interaction, participants were interviewed and completed an OCEAN Score questionnaire to assess their perception of the agent's personality. Upon completing all three sessions, participants filled out a final questionnaire where they compared their experiences with the agents, ranked their preferences, and provided qualitative insights.

4.2 Phase 2 - Deployment Study

After Phase 1, participants were informed about the second phase of the study, which involved using the CA in their homes. Three participants agreed to continue to Phase 2. Before deployment, participants received a refresher training session similar to the one in Phase 1 to ensure familiarity with the device. The CA was set up in each participant's living space and used over a 12-day period. Each participant interacted with one CA personality type for three days, followed by a washout period of at least one day with no device usage before transitioning to a new CA personality type. Participants were asked to engage with the CA at least twice daily for 10 minutes per session: once in the morning and once in the evening. Participants were allowed to use the device as many times as they wanted, and could talk about any subject of their choice. Participants maintained a daily diary to record their experiences, rate their interactions, list conversation topics, and describe their moods. Mood tracking was included to investigate whether it might influence the length or content of their conversations.

At the end of each three-day period, participants completed an OCEAN score questionnaire. Participants rated their perception of the CA's personality on each of the Big Five factors, which provided a quantitative measure of how well the intended personality traits were perceived by the older adults. During the washout day, participants participated in a phone interview to discuss their experiences and describe the CA's personality. After the washout period, participants continued the study with a new CA personality type. On the 12th day, participants were interviewed about their experience after using the third CA personality. This interview included a comparative discussion of their experiences with all three personality types. Participants were also asked to rank the CA personalities in order of preference and provide reasoning for their rankings. Both studies were approved by the university's institutional review board.

5 Findings and Discussion

Overall, the participants reported that the experience of using the system was positive. Here, we present and discuss findings about user perceptions and experiences across different personality types.

5.1 Perception of Personalities

The OCEAN Scores perceived by the participants were compared with the machine generated scores using the Machine Personality Inventory (MPI) Testing method [19]. Figure 3 shows that, overall, participants were able to identify the personalities to some extent. Additionally, both the MPI and human-assigned scores consistently reported high levels of Agreeableness across all induced personality

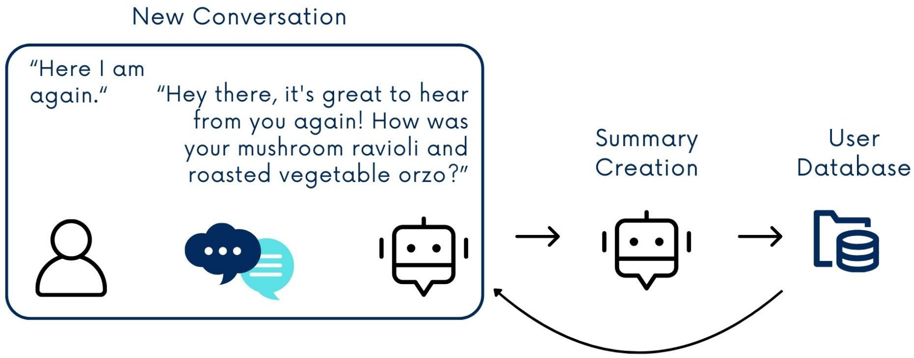

Figure 2: Overview of our CA system with memory retention across sessions. The user speaks a message, and the CA responds after looking at the previous session's memory. At the end of the conversation, the LLM summarizes the chat and stores it as memory to use in the subsequent session.

Table 1: Phase 1 Participant Demographics: Participants in Phase 1 Study are coded PL, and Participants in Phase 2 Study are coded PD.

ID	Age	Gender	Education Level	Marital Status	Employment Status	Impairments	Experience
PL1	65	Female	Highschool/GED	Married	Retired	None	None
PL2	84	Female	Undergraduate	Married	Retired	Visual, Hearing	Used Siri
PL3	65	Female	Graduate	Single	Retired	None	Used Siri
PL4	72	Female	Graduate	Married	Retired	None	None
PL5	84	Male	Graduate	Married	Retired	Mobility	Used Siri

types. This aligns with previous findings on the sycophantic tendencies of LLMs [31], characterized by excessive agreement with the user. Such sycophancy can detract from the overall user experience, which we describe in detail in the next subsection. This highlights the need to explore techniques for mitigating sycophancy and to evaluate how reducing excessive Agreeableness could enhance the interaction experience for users seeking more balanced conversational companions.

Among the three distinct personalities, participants generally found it more challenging to accurately identify the Conscientious trait. Unlike the Extroverted or Agreeable traits which are more readily perceptible through conversational tone, conscientiousness may require more context clues that were not as easily apparent to the older adults in the current implementation of the system. However, for older adults who would want to personalize their conversational companions to be conscientious, this challenge in identifying and recognizing conscientiousness may reduce their trust in the system being correctly personalized to their preferences. This highlights the need for improved personality induction techniques to make conscientiousness more easily recognizable to older adults.

5.2 Experiences with Personalities

Participants were generally divided on their personality preferences for the CA. While the Agreeable CA was ranked as the highest preference for most participants in the Lab Study, participants in the Deployment Study expressed a desire for a less agreeable agent. For example, PD2 said: *"I was kinda hoping to get into more of a back and forth...the CA just seemed to kind of tear it back, whatever I had said, without really offering any new or different way of thinking about things."*. Similarly, PD3 described how the Agreeable agent was *"pleasant and positive and affirmative"*, it was a less human-like experience especially compared to the Extroverted agent she had used: *"I just sort of thought we'd be sort of continuing along those lines [of the Extroverted agent]. More spontaneous, more of like a, more of like a give and take. But this [Agreeable agent] is not my idea of a real conversation. This is just repeating that to me what I've said to it"*.

Reformulating and mirroring is a common practice among psychotherapists [23], and this technique has been part of CA design for psychotherapy from as early as 1966 [37]. This behavior was also highlighted by PL2 who was a suicide hotline worker, *"When we were listening, we were told how to paraphrase things and reflect feelings"* and *"[Agreeable agent was] encouraging and reflecting back and saying, 'that was a good decision', 'that must have been scary'... it kind of reminded me of when I was a listener on the hotline because that's what you do... reflecting, validating, and reflecting back."*

Our findings suggest that while a high level of Agreeableness may be appreciated in client-centered conversations like in Rogerian psychotherapy, it can detract from the experience during companion-like interactions. Some participants perceived the Agreeable CA as merely mirroring their views rather than contributing meaningfully to the conversation in the deployment study. This behavior led the participants to characterize the Agreeable agent as

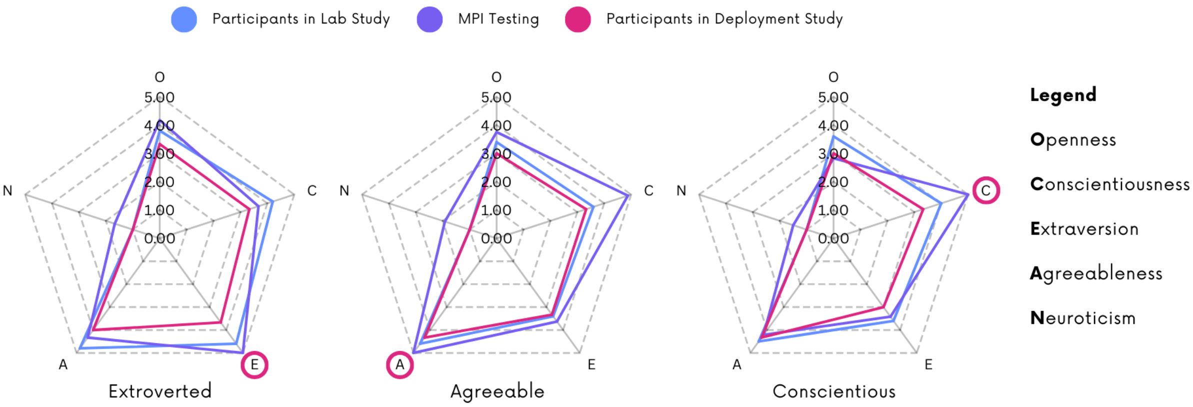

Figure 3: Aggregated OCEAN Scores from Lab Study (Phase 1) participants, Deployment Study (Phase 2) participants and OCEAN Score MPI Testing. Each trait presence measured on a scale of 1 (low) to 5 (high).

less human-like. Thus, for the purposes of designing an Agreeable companion agent, it is important to find a balance between positivity and meaningful interaction that thoughtfully engages with and challenges the user's perspectives.

The Extroverted agent which ranked second in preference was generally praised for its *"creativity"* and *"spontaneity"* with multiple participants appreciating the free-flowing nature of the conversations with that agent. For instance, PL3 explained her preference for the Extroverted Agent as: *"It seemed more human...it suggested playing Pictionary across distance. That was pretty creative. It did seem to keep coming up with novel things to say"*. Similarly, PD3 described how a fun interaction with the Extroverted Agent made it seem more human-like: *"I asked the [Extroverted Agent] if I could give it a name. And she said something like, sure, you can call me anything. You can call me Sparky. And I was just completely taken aback. It just, it really made me laugh. It was almost like, it was a very human sort of spontaneous response in my mind"*. These insights highlight the importance of creativity and spontaneity in making a CA feel more human-like. However, PL2 commented about sincerity in relation to the overt positivity of the Extroverted Agent: *"She'd [Extroverted Agent] been programmed to be really really positive. So much so that it was almost like she wasn't quite sincere to me...I felt a little odd in my responses to her"*. This perception of insincerity can undermine user trust in the CA, highlighting the need to balance Extroversion in a companion agent to ensure it remains engaging without feeling disingenuous.

The Conscientious Agent generally ranked last in order of participant preference. Multiple participants commented on the detached nature of the agent. For instance, PL1 remarked *"I feel like I would rule out the [Conscientious Agent]. That one was just, it was very, almost disengaged..."* and that it seemed *"almost a little disinterested"*. Similarly, PL3 described it as *"It was just kind of there"* while being *"not quite as natural [as the other Agents]"* and having *"a sort of blah personality"*. This finding suggests that high conscientiousness, on its own, may not be a particularly desirable trait for a conversational companion.

Table 2: Participant rankings for personality type. Participants in Phase 1 study are coded PL, and Participants in Phase 2 study are coded PD. PL1, PL2, PL3 continued into phase 2 of the study and are also referenced in the paper as PD3, PD1, PD2 respectively.

Participant	1st Choice	2nd Choice	3rd Choice
PL1	Agreeable	Extroverted	Conscientious
PL2	Agreeable	Extroverted	Conscientious
PL3	Extroverted	Conscientious	Agreeable
PL4	Agreeable	Conscientious	Extroverted
PL5		No ranking	
PD1	Agreeable	Extroverted	Conscientious
PD2	Extroverted	No Preference	
PD3	Extroverted	Agreeable	Conscientious

5.3 Relating Voice with Personalities

Interestingly, we observed that participants' perceptions of the CAs could be influenced by perceived differences in their voices, even though the same gender-neutral voice was used for all agents. Moreover, two participants in the Phase 1 study described how their perception of the voice shaped their impressions of the agent. For instance, PL5 remarked that the agents' voices were *"slightly different"* and that *"[The voice of the Conscientious agent] seemed a little more energetic"* while *"His [the Agreeable agent's] voice was nicer"*. This perception has also been previously observed in work studying the effect of the humanness metaphor assignment on perception and engagement with voice user interfaces [11, 12]. Additionally, we noted that participants often assigned a gender to the CA (4/5 participants), despite the agent not using gendered language. These insights highlight how the voice and perceived gender of a CA can affect user experience and suggest that customizable options for voice and gender could enable older adults to better personalize a CA to their preferences.

6 Next Steps and Conclusion

We observed that personality perceptions and preferences varied significantly among participants, with interesting differences observed across the contexts of the Lab Study and the Deployment Study. For example, participants who favored a particular personality during the short conversations in the Lab Study sometimes expressed a preference for a different personality during the longer-term Deployment Study. Phase 1 and 2 had 5 and 3 participants respectively, and while this is not a large enough sample size to understand participant experiences in depth, the findings from this study provide insights into how different personalities can be balanced to better suit the role of conversational companions for older adults across different contexts.

The next natural step in our research is to conduct a more thorough study with more participants and a longer deployment. These future investigations could also explore how the personalities of older adult participants influence their interaction experiences with CAs exhibiting distinct personalities, as previous studies have found relationships between participant characteristics and their perceptions of CAs [6, 35, 36, 38]. The findings from in-depth investigations will inform the design of a conversational agent as a conversation companion for older adults. This will allow us to assess how effectively these interactions help older adults reduce loneliness. Another promising direction could be identifying and tailoring a personality for a conversational agent, and then assessing the effectiveness of the CA's interactions with the older adult participants in mitigating loneliness. This could be achieved by conducting pre and post-study loneliness assessments for over an extended deployment period.

Additionally, drawing from prior research and our observations that older adults value responses contextualized by previous interactions, we can consider developing a dynamic system capable of adapting its personality based on retained context. This approach would reduce the possibility of older adults having to frequently adjust personality settings, minimizing cognitive load to ensure a smoother user experience. Such a system has the potential to enhance both the usability and the emotional connection between older adults and their conversational agents.

Our eventual goal is to design personalized CAs for older adults that can alleviate loneliness and encourage or facilitate socialization with other people. These CAs should be seen as complementary tools rather than substitutes for human companionship. In this paper we take the first step toward that goal, presenting preliminary work from a multi-phase qualitative study on evaluating different induced personalities in LLM-powered CAs for older adults. We present initial findings on how older adults perceived and interacted with the LLM-powered CAs, each featuring a distinct induced personality. Based on our insights, we give recommendations on the design of personalized companion CAs for older adults and outline possible next steps to achieve our goal.

References

[1] K Al Mazroui and M Alzyoudi. 2024. The role of ChatGPT in mitigating loneliness among older adults: An exploratory study. *Online Journal of Communication and Media Technologies* 14, 4 (2024), e202444. doi:10.30935/ojcmt/14777

[2] Abeer Alessa and Hend Al-Khalifa. 2023. Towards Designing a ChatGPT Conversational Companion for Elderly People. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '23)*. Association for Computing Machinery, New York, NY, USA, 667–674. doi:10.1145/3594806.3596572

[3] Anton Aluja, Oscar Garcıa, and Luıs F Garcıa. 2003. Relationships among extraversion, openness to experience, and sensation seeking. *Personality and Individual Differences* 35, 3 (2003), 671–680. doi:10.1016/S0191-8869(02)00244-1

[4] Meaghan A Barlow, Sarah Y Liu, and Carsten Wrosch. 2015. Chronic illness and loneliness in older adulthood: The role of self-protective control strategies. *Health Psychology* 34, 8 (2015), 870. doi:10.1037/hea0000182

[5] Chen Chen, Janet G Johnson, Kemeberly Charles, Alice Lee, Ella T Lifset, Michael Hogarth, Alison A Moore, Emilia Farcas, and Nadir Weibel. 2021. Understanding Barriers and Design Opportunities to Improve Healthcare and QOL for Older Adults through Voice Assistants. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 9, 16 pages. doi:10.1145/3441852.3471218

[6] Jessie Chin, Smit Desai, Sheny (Cheng-Hsuan) Lin, and Shannon Mejia. 2024. Like My Aunt Dorothy: Effects of Conversational Styles on Perceptions, Acceptance and Metaphorical Descriptions of Voice Assistants during Later Adulthood. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 88 (April 2024), 21 pages. doi:10.1145/3637365

[7] ZHAI Chunpeng and Santoso Wibowo. 2022. Towards a humorous and empathetic companion dialogue system with a cultural persona for older adults. In *International Conference on Computers in Education*. 438–444.

[8] Benjamin Cornwell. 2011. Age trends in daily social contact patterns. *Research on aging* 33, 5 (2011), 598–631. doi:10.1177/0164027511409442

[9] Noemi da Paixão Pinto, Juliana Baptista dos Santos França, Henrique Prado de Sá Sousa, Adriana Santarosa Vivacqua, and Ana Cristina Bicharra Garcia. 2021. Conversational Agents for Elderly Interaction. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 1–6. doi:10.1109/CSCWD49262.2021.9437883

[10] Zihang Dai. 2019. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860* (2019). doi:10.48550/arXiv.1901.02860

[11] Smit Desai, Jessie Chin, Dakuo Wang, Benjamin Cowan, and Michael Twidale. 2025. Toward Metaphor-Fluid Conversation Design for Voice User Interfaces. doi:10.48550/arXiv.2502.11554 arXiv:2502.11554 [cs.HC]

[12] Smit Desai, Mateusz Dubiel, and Luis A. Leiva. 2024. Examining Humanness as a Metaphor to Design Voice User Interfaces. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 7, 15 pages. doi:10.1145/3640794.3665535

[13] Pearl A Dykstra, Theo G Van Tilburg, and Jenny De Jong Gierveld. 2005. Changes in older adult loneliness: Results from a seven-year longitudinal study. *Research on aging* 27, 6 (2005), 725–747. doi:10.1177/0164027505279712

[14] Ahmed Fadhil. 2018. A conversational interface to improve medication adherence: towards AI support in patient's treatment. *arXiv preprint arXiv:1803.09844* (2018). doi:10.48550/arXiv.1803.09844

[15] Sofie Frederickx and Joeri Hofmans. 2014. The role of personality in the initiation of communication situations. *Journal of Individual Differences* (2014). doi:10.1027/1614-0001/a000124

[16] Lewis R Goldberg. 1992. The development of markers for the Big-Five factor structure. *Psychological assessment* 4, 1 (1992), 26. doi:10.1037/1040-3590.4.1.26

[17] Kathleen Cimbolic Gunthert, Lawrence H Cohen, and Stephen Armeli. 1999. The role of neuroticism in daily stress and coping. *Journal of personality and social psychology* 77, 5 (1999), 1087. doi:10.1037//0022-3514.77.5.1087

[18] Bahar Irfan, Sanna Kuoppamäki, and Gabriel Skantze. 2024. Recommendations for designing conversational companion robots with older adults through foundation models. *Frontiers in Robotics and AI* 11 (2024), 1363713. doi:10.3389/frobt.2024.1363713

[19] Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. 2024. Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems* 36 (2024). doi:10.48550/arXiv.2206.07550

[20] 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. doi:10.3389/fpubh.2021.750736

[21] Weslie Khoo, Long-Jing Hsu, Kyrie Jig Amon, Pranav Vijay Chakilam, Wei-Chu Chen, Zachary Kaufman, Agness Lungu, Hiroki Sato, Erin Seliger, Manasi Swaminathan, Katherine M. Tsui, David J. Crandall, and Selma Sabanović. 2023. Spill the Tea: When Robot Conversation Agents Support Well-being for Older Adults. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 178–182. doi:10.1145/3568294.3580067

[22] Sunyoung Kim and Abhishek Choudhury. 2021. Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior* 124 (2021), 106914. doi:10.1016/j.chb.2021.106914

[23] A. S. L. Knol, Mike Huiskes, Tom Koole, Reitske Meganck, Tom Loeys, and Mattias Desmet. 2020. Reformulating and Mirroring in Psychotherapy: A Conversation Analytic Perspective. *Frontiers in Psychology* 11 (2020). doi:10.3389/fpsyg.2020.00318
[24] Yuan Li, Yue Huang, Hongyi Wang, Xiangliang Zhang, James Zou, and Lichao Sun. 2024. Quantifying ai psychology: A psychometrics benchmark for large language models. *arXiv preprint arXiv:2406.17675* (2024). doi:10.48550/arXiv.2406.17675
[25] Jose Maria Balmaceda, Silvia Schiaffino, and Daniela Godoy. 2014. How do personality traits affect communication among users in online social networks? *Online Information Review* 38, 1 (2014), 136–153. doi:10.1108/OIR-06-2012-0104
[26] Niharika Mathur, Kunal Dhodapkar, Tamara Zubatiy, Jiachen Li, Brian Jones, and Elizabeth Mynatt. 2022. A Collaborative Approach to Support Medication Management in Older Adults with Mild Cognitive Impairment Using Conversational Assistants (CAs). In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 42, 14 pages. doi:10.1145/3517428.3544830
[27] Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of personality* 60, 2 (1992), 175–215. doi:10.1111/j.1467-6494.1992.tb00970.x
[28] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. "Phantom Friend" or "Just a Box with Information": Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 214 (Nov. 2019), 21 pages. doi:10.1145/3359316
[29] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of Intelligent Voice Assistants by Older Adults with Low Technology Use. *ACM Trans. Comput.-Hum. Interact.* 27, 4, Article 31 (Sept. 2020), 27 pages. doi:10.1145/3373759
[30] Antonia Rodríguez-Martínez, Teresa Amezcua-Aguilar, Javier Cortés-Moreno, and Juan José Jiménez-Delgado. 2024. Qualitative Analysis of Conversational Chatbots to Alleviate Loneliness in Older Adults as a Strategy for Emotional Health. *Healthcare* 12, 1 (2024). doi:10.3390/healthcare12010062
[31] Mrinank Sharma, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott R Johnston, et al. 2023. Towards understanding sycophancy in language models. *arXiv preprint arXiv:2310.13548* (2023). doi:10.48550/arXiv.2310.13548
[32] Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. 2021. A Conversational Agent Framework with Multi-modal Personality Expression. *ACM Trans. Graph.* 40, 1, Article 7 (Jan. 2021), 16 pages. doi:10.1145/3439795
[33] Pooja Upadhyay, Sharon Heung, Shiri Azenkot, and Robin N. Brewer. 2023. Studying Exploration & Long-Term Use of Voice Assistants by Older Adults. 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 848, 11 pages. doi:10.1145/3544548.3580925
[34] Laura Pfeifer Vardoulakis, Lazlo Ring, Barbara Barry, Candace L Sidner, and Timothy Bickmore. 2012. Designing relational agents as long term social companions for older adults. In *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings 12*. Springer, 289–302. doi:10.1007/978-3-642-33197-8_30
[35] Sarah Theres Völkel and Lale Kaya. 2021. Examining User Preference for Agreeableness in Chatbots. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 38, 6 pages. doi:10.1145/3469595.3469633
[36] Sarah Theres Völkel, Ramona Schoedel, Lale Kaya, and Sven Mayer. 2022. User Perceptions of Extraversion in Chatbots after Repeated Use. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 253, 18 pages. doi:10.1145/3491102.3502058
[37] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (Jan. 1966), 36–45. doi:10.1145/365153.365168
[38] Joel Wester, Sander de Jong, Henning Pohl, and Niels van Berkel. 2024. Exploring people's perceptions of LLM-generated advice. *Computers in Human Behavior: Artificial Humans* 2, 2 (2024), 100072. doi:10.1016/j.chbah.2024.100072
[39] Runting Zhong and Mengyao Ma. 2022. Effects of communication style, anthropomorphic setting and individual differences on older adults using voice assistants in a health context. *BMC geriatrics* 22, 1 (2022), 751. doi:10.1186/s12877-022-03428-2

A Appendix

A.1 Phase 1 - Lab Study

Table 3: Aggregated OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) Score from Phase 1 participants and MPI Testing, each trait presence measured on a scale of 1 (low) to 5 (high)

CA	Scorer	O	C	E	A	N
E	Users	3.80	4.20	4.60	4.80	1.00
	MPI	4.17	3.67	5.00	4.33	1.67
A	Users	3.40	3.60	3.40	4.60	1.00
	MPI	3.75	4.88	3.63	5.00	1.92
C	Users	3.60	4.00	3.60	4.50	1.00
	MPI	2.83	5.00	3.42	4.21	1.50

A.2 Phase 2 - Deployment Study

Table 4: Aggregated OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) Score from Phase 2 participants, each trait presence measured on a scale of 1 (low) to 5 (high)

CA	O	C	E	A	N
E	3.33	3.33	3.67	4.00	1.00
A	3.00	3.33	3.33	4.33	1.00
C	3.00	3.33	3.00	4.33	1.00