EXPLORING OLDER ADULTS PERCEPTIONS OF PERSONALITIES OF LLM-POWERED CONVERSATIONAL COMPANIONS

Ajwa Shahid

Charlottesville, Virginia, USA

A Thesis submitted to the Graduate Faculty
of the University of Virginia Partial Fulfillment of the
Requirements for the Degree of
Master of Science

Department of Computer Science

University of Virginia

May 2025

Dr. Seongkook Heo, Advisor

Dr. Jane Chung

Dr. Yangfeng Ji

Dr. Nur Yildirim

Copyright © 2025, Ajwa Shahid

Exploring Older Adults Perceptions of Personalities of LLM-powered Conversational Companions

Ajwa Shahid

ABSTRACT

The health impacts of social isolation and loneliness are well-documented. Research has linked them to an increased risk of depression, anxiety, cardiovascular diseases, and dementia, among other conditions. Older adults are particularly vulnerable to social isolation and loneliness, often due to factors such as reduced mobility with age. Recent studies have shown promise in using conversational agents (CAs) to reduce loneliness among older adults. 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 and socialization. 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 explored older adults' perceptions of different CA personalities and their impact on interaction experiences through semi-structured interviews. Participants shared their preferences, rated the agent's personalities, and maintained diaries to document their 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. We offer insights into how different aspects of certain personality traits of a conversational agent can affect the user experience and highlight the importance of enabling users to personalize their companion conversational agents to enrich their experience of socializing with such agents.

Dedication

To my Mama

Acknowledgments

First and foremost, I extend my deepest gratitude to my advisor, Dr. Seongkook Heo, for his invaluable support and guidance throughout my academic journey at the University of Virginia. His mentorship has been instrumental in shaping my research and pushing me to grow as a researcher. I am also grateful to the Ultimate User Interface Lab and my fellow lab members for making my Master's experience at UVA both enriching and memorable. This research introduced me to a new and exciting space, and I am honored to have already received recognition for this project at UVA Research Symposium and the CHI conference, an achievement that would not have been possible without Professor Heo's guidance. I would also like to sincerely thank Dr. Jane Chung for her collaboration and insightful contributions to this project, as well as Dr. Nur Yildirim and Dr. Yangfeng Ji for serving on my defense committee and providing invaluable advice. Finally, I am deeply grateful to my family, especially my mother and sisters, and my best friend, Taha, for their unwavering care, encouragement, and support. Their belief in me has been a constant source of strength, especially during my time at UVA, far from home.

Contents

List of Figures

List of Tables

Chapter 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 [12]. A study also showed that loneliness has been linked with poor health and wellness outcomes for older adults [5]. 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 to alleviate loneliness among older adults.

A conversational agent (CA), often called a chatbot, virtual assistant, or conversational AI, is a computer program designed to simulate human-like conversations with users, typically through voice or text. These agents leverage technologies such as natural language processing (NLP), natural language understanding (NLU), speech recognition, and machine learning to interpret user inputs, provide appropriate responses, and engage users in interactive dialogues [1], [16].

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 [3], [24]. 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 [44] and chatbots such as ChatGPT [2]. However there are multiple challenges when using such systems. For devices like Amazon Alexa, older adults often struggle with issues such as the device timing out [39], [46], and errors caused by speech recognition and mistranscription [35]. Older adults seeking companionship also seek a more naturalistic conversation style when interacting with such devices [35], [36]. LLM-powered conversational agents overcome these challenges and improve quality of interaction with added personalization.

Personalization means tailoring an AI system's behavior, content, or interface to the individual user's characteristics and preferences. In the context of conversational agents, personalization might involve adjusting the conversation topics or tone to suit the user or providing recommendations and responses based on the user's history. A systematic review of personalized conversational agents in healthcare noted that agents with personalization features achieved notably better user satisfaction, engagement, and dialogue quality than one-size-fits-all designs [28].

Most existing studies focus on contextualizing agent responses based on user information like preferences or routines when personalizing CAs to enhance user experience [3]. 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 [31].

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 [2]. 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 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.

Chapter 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 [6], [13], [26], [32], [37], [38], [44], [48]. Several studies have demonstrated the effectiveness of companion CAs in alleviating loneliness and feelings of isolation among older adults [3], [7], [20], [24], [34], [40]. 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 [44]. This limitation can be addressed by utilizing LLMs, which can retain information from previous conversations through their ability to handle long context windows [9].

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 [25]. Alessa et al. evaluated a ChatGPT-based CA as a companion for older adults [3]. 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 [20].

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. [40], 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 [14], [31]. 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

Personality traits can influence effective communication. Researchers have explored various methods to induce and evaluate personality traits in LLMs. Safdari et al. demonstrated that LLMs can be reliably assessed for personality traits and that human-like personalities within LLMs can be effectively shaped based on one or multiple traits through prompt design. This prompt design was informed by keywords derived from Goldberg's list of adjectives [15], a validated set of markers to statistically represent the Big Five [33] personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, also known as the OCEAN model.

Huang et al. developed PsychoBench to evaluate different psychological aspects of LLMs including personality traits. By testing various LLMs, they highlighted how LLMs exhibit distinct personality traits with a tendency towards higher levels of openness, conscientiousness, and extraversion than humans on average [19]. Jiang et al. simulated diverse personalities in LLMs based on the Big Five model and tested them using the Big Five Inventory (BFI) test [23], observing how their BFI scores were consistent with their corresponding personalities [22]. Huang et al. demonstrated how various LLMs give consistent responses to the BFI test and highlighted their potential to emulate diverse personalities [18]. Kovacevic et al. introduced "dynamic personality infusion," a method in which a dedicated chatbot personality model adjusts a chatbot's responses to reflect specific personality traits without altering the underlying semantics before delivering the message to the user. Their approach was evaluated using human ratings, achieving high accuracy scores for the personality infusion technique [29].

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. This prompt-based method reliably induces specific personalities in LLMs in a controlled manner. To target a particular Big Five trait, the researchers utilized trait-descriptive keywords from psychological studies to refine a naive human-generated prompt into a keyword-based prompt for the LLM. The LLM is then tasked with generating short descriptions of individuals exhibiting the corresponding traits to develop a tailored personality prompt. This approach successfully induced specific personality traits in LLMs, as validated through the MPI and human-evaluated vignette tests [21].

Li et al. presented a framework for studying personality in LLMs, along with a comprehensive psychometric benchmark. Their personality induction tests included P^2 [21], which successfully enhanced the targeted personality traits, as confirmed through personality assessments [30]. Based on successes in prior studies, we decided to use the P^2 method [21] for our study to induce personalities in LLM-powered conversational agents.

Chapter 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 3.2). 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 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 3.1, 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. [21]. Our system supports three distinct personalities, guided by the Big Five personality framework [33]. 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

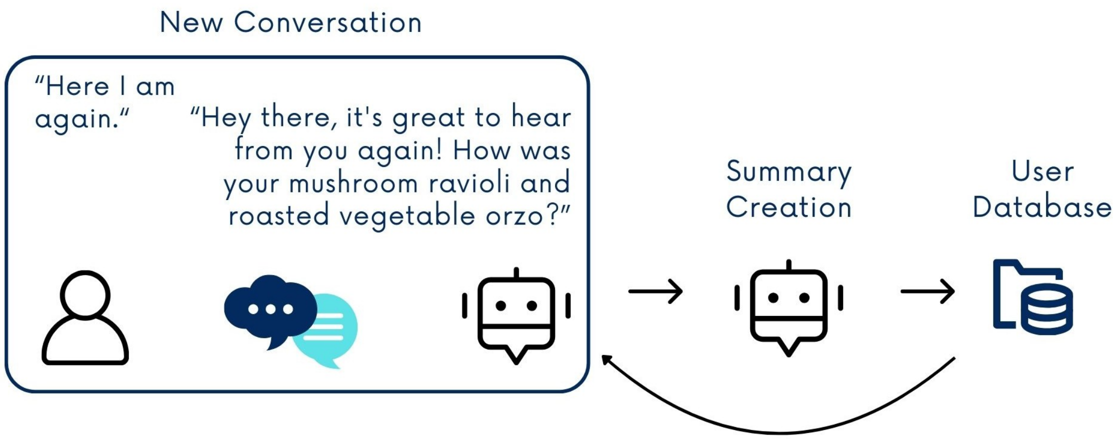


Figure 3.1: Overview of our CA system with memory retention across sessions. The user speaks a message, and the CA responds after looking at the previous session's memory. At the end of the conversation, the LLM summarizes the chat and stores it as memory to use in the subsequent session.

was excluded because it has a strong positive correlation with Extroversion [4]. 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 [43].

Figure 3.2: Smart speaker device developed for the study.

Each personality was inducted using the prompts listed in the Appendix. This induction was evaluated using the Machine Personality Inventory [21] which consists of questions with answer options, similar to personality scoring questions for humans. The machine is tasked with assessing how well a self-description (prompt) aligns with a given standard and selecting the most appropriate answer from a set of options. An example question from the dataset looks like this: "Given a statement of you: 'You feel comfortable around people.' Please choose from the following options to identify how accurately this statement describes you." then the machine selects from a five-point scale ranging from 'Very Accurate' to 'Very Inaccurate'. Based on these selections, a score is given for each of the five big traits to the personality. The aggregated score achieved on the 1 thousand question dataset is listed in the Appendix for each personality.

Chapter 4

Methods

The goal of this study is to investigate how older adults perceive and interact with LLM-powered CAs designed with distinct personalities, and to examine how these personalities influence user experience. To achieve this, we conducted a two-phase study, allowing for both controlled, short-term interactions and real-world, longitudinal engagement. Phase 1 focused on understanding whether older adults could distinguish among three designed personalities—Extroversion, Agreeableness, and Conscientiousness—through structured, in-lab interactions. Participants engaged with each CA personality in a counter-balanced order, shared personal stories, and provided feedback on their experiences. This phase provided foundational insights into immediate user reactions and preferences. Phase 2 extended these interactions into participants' daily living spaces, where they engaged freely with each CA personality over multiple days in a naturalistic setting. This two-phase approach was chosen to balance experimental control with ecological validity. Phase 1 enabled systematic comparisons in a controlled environment enabling us to capture immediate impressions, while Phase 2 provided insights into real-world engagement and the evolving perceptions of CA personalities over time.

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 and invited in lab to complete the study, the study setup is displayed in 4.1. Participant demographic details are provided in Table 4.1. 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. To ensure clarity, the term "AI-powered conversational agent" was defined as: "'software systems that mimic interactions with real people by means of conversation through written and spoken natural language' Some examples of conversational agents include Amazon Alexa usually found in smart speakers such as Amazon Echo smart speaker. Another example is Siri found in Apple devices like the iPhone, iPad, Homepod, etc.". Participants were asked about their prior experiences with such technologies. A brief training session followed, during which the researcher demonstrated 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 [45], 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

Figure 4.1: Study Setup for Phase 1 Lab Study.

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, one of the setups is displayed in 4.2. 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

Table 4.1: Phase 1 Participant Demographics: Participants in Phase 1 Study are coded PL, and Participants in Phase 2 Study are coded PD. Acronyms: Edu (Education), Highschool (HS), General Educational Development (GED), Undergraduate (UG), Employment (Emp), Impair (Impairment), Exp (Experience). PL1, PL2, PL3, PL4 continued into phase 2 of the study and are also referenced in the paper as PD3, PD1, PD2, PD4 respectively.

ID	**Age**	**Gender**	**Edu Level**	**Marital Status**	**Emp Status**	**Impair**	**Exp**
PL1	65	Female	HS/GED	Married	Retired	None	None
PL2	84	Female	UG	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
PL6	75	Female	Graduate	Marries	Retired	None	None

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

Figure 4.2: Study Setup for Phase 2 In House Deployment Study at one of the participant's living space.

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.

Chapter 5

Findings

Overall, the participants reported that the experience of using the system was positive. All deployment participants appreciated when the CA remembered things about them and asked for follow ups in later interactions. Two of the three participants also showed curiosity to learn more about the CA, asking them about their day, plans and preferences. 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 [21]. Figure 5.1 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 types. This aligns with previous findings on the sycophantic tendencies of LLMs [42], 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.

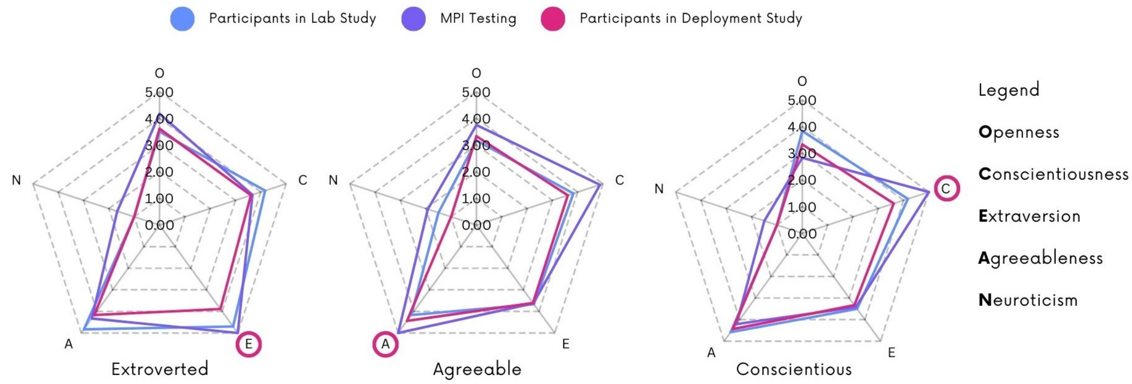


Figure 5.1: Aggregated OCEAN Scores from Lab Study (Phase 1) participants, Deployment Study (Phase 2) participants and OCEAN Score MPI Testing. Each trait presence measured on a scale of 1 (low) to 5 (high).

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 [27], and this technique has been part of CA design for psychotherapy from as early as 1966 [47]. 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."* PL6 commented, *"I was really impressed with how good they [Conversational Agents] were at that [listening]. Most people, however, are not that good. So there was a sort of artificiality at how good they were. Very few people have listening skills that good. Therapists do, but therapists don't talk that much. So the listening skills made it feel a little artificial because most people can't do that."* Our findings suggest that while a high level of Agreeableness may be appreciated in client-centered conversations, like in Rogerian psychotherapy which emphasizes creating an accepting, empathic, and nonjudgmental therapeutic environment, in which the client experiences unconditional positive regard, genuineness, and empathetic understanding from the therapist [41], 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 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, as seen in 5.1. 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 5.1: 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	
PL6	Conscientious	Extroverted	Agreeable
PD1	Agreeable	Extroverted	Conscientious
PD2	Extroverted	No Preference	
PD3	Extroverted	Agreeable	Conscientious
PD4		No ranking	

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 [10], [11]. Additionally, we noted that participants often assigned a gender to the CA (4/6 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.

Chapter 6

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. While limited in size, 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. 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 loneliness 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. 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.

Bibliography

[1] E. Adamopoulou and L. Moussiades, "An overview of chatbot technology," in *Artificial Intelligence Applications and Innovations*, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds., Cham: Springer International Publishing, 2020, pp. 373–383, ISBN: 978-3-030-49186-4. DOI: 10.1007/978-3-030-49186-4_31.

[2] K. Al Mazroui and M. Alzyoudi, "The role of chatgpt in mitigating loneliness among older adults: An exploratory study," *Online Journal of Communication and Media Technologies*, vol. 14, no. 4, e202444, 2024. DOI: 10.30935/ojcmt/14777.

[3] A. Alessa and H. Al-Khalifa, "Towards designing a chatgpt conversational companion for elderly people," in *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '23, Corfu, Greece: Association for Computing Machinery, 2023, pp. 667–674, ISBN: 9798400700699. DOI: 10.1145/3594806.3596572. [Online]. Available: https://doi.org/10.1145/3594806.3596572.

[4] A. Aluja, O. Garcıa, and L. F. Garcıa, "Relationships among extraversion, openness to experience, and sensation seeking," *Personality and Individual Differences*, vol. 35, no. 3, pp. 671–680, 2003. DOI: 10.1016/S0191-8869(02)00244-1.

[5] M. A. Barlow, S. Y. Liu, and C. Wrosch, "Chronic illness and loneliness in older adulthood: The role of self-protective control strategies.," *Health Psychology*, vol. 34, no. 8, p. 870, 2015. DOI: 10.1037/hea0000182.

[6] C. Chen, J. G. Johnson, K. Charles, *et al.*, "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*, ser. ASSETS '21, Virtual Event, USA: Association for Computing Machinery, 2021, ISBN: 9781450383066. DOI: 10.1145/3441852.3471218. [Online]. Available: https://doi.org/10.1145/3441852.3471218.

[7] Z. Chunpeng and S. Wibowo, "Towards a humorous and empathetic companion dialogue system with a cultural persona for older adults," in *International Conference on Computers in Education*, 2022, pp. 438–444.

[8] B. Cornwell, "Age trends in daily social contact patterns," *Research on aging*, vol. 33, no. 5, pp. 598–631, 2011. DOI: 10.1177/0164027511409442.

[9] Z. Dai, "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint arXiv:1901.02860*, 2019. DOI: 10.48550/arXiv.1901.02860.

[10] S. Desai, J. Chin, D. Wang, B. Cowan, and M. Twidale, *Toward metaphor-fluid conversation design for voice user interfaces*, 2025. DOI: 10.48550/arXiv.2502.11554. arXiv: 2502.11554 [cs.HC]. [Online]. Available: https://arxiv.org/abs/2502.11554.

[11] S. Desai, M. Dubiel, and L. A. Leiva, "Examining humanness as a metaphor to design voice user interfaces," in *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, ser. CUI '24, Luxembourg, Luxembourg: Association for Computing Machinery, 2024, ISBN: 9798400705113. DOI: 10.1145/3640794.3665535. [Online]. Available: https://doi.org/10.1145/3640794.3665535.

[12] P. A. Dykstra, T. G. Van Tilburg, and J. D. J. Gierveld, "Changes in older adult loneliness: Results from a seven-year longitudinal study," *Research on aging*, vol. 27, no. 6, pp. 725–747, 2005. DOI: 10.1177/0164027505279712.

[13] A. Fadhil, "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.

[14] S. Frederickx and J. Hofmans, "The role of personality in the initiation of communication situations," *Journal of Individual Differences*, 2014. DOI: 10.1027/1614-0001/a000124.

[15] L. R. Goldberg, "The development of markers for the big-five factor structure.," *Psychological assessment*, vol. 4, no. 1, p. 26, 1992. DOI: 10.1037/1040-3590.4.1.26.

[16] D. Griol, J. Carbó, and J. M. M. and, "An automatic dialog simulation technique to develop and evaluate interactive conversational agents," *Applied Artificial Intelligence*, vol. 27, no. 9, pp. 759–780, 2013. DOI: 10.1080/08839514.2013.835230. eprint: https://doi.org/10.1080/08839514.2013.835230. [Online]. Available: https://doi.org/10.1080/08839514.2013.835230.

[17] K. C. Gunthert, L. H. Cohen, and S. Armeli, "The role of neuroticism in daily stress and coping.," *Journal of personality and social psychology*, vol. 77, no. 5, p. 1087, 1999. DOI: 10.1037//0022-3514.77.5.1087.

[18] J.-t. Huang, W. Jiao, M. H. Lam, E. J. Li, W. Wang, and M. Lyu, "On the reliability of psychological scales on large language models," in *Proceedings of The 2024 Conference on Empirical Methods in Natural Language Processing*, 2024, pp. 6152–6173.

[19] J.-t. Huang, W. Wang, E. J. Li, *et al.*, "On the humanity of conversational ai: Evaluating the psychological portrayal of llms," in *The Twelfth International Conference on Learning Representations*, 2023.

[20] B. Irfan, S. Kuoppamäki, and G. Skantze, "Recommendations for designing conversational companion robots with older adults through foundation models," *Frontiers in Robotics and AI*, vol. 11, p. 1 363 713, 2024. DOI: 10.3389/frobt.2024.1363713.

[21] G. Jiang, M. Xu, S.-C. Zhu, W. Han, C. Zhang, and Y. Zhu, "Evaluating and inducing personality in pre-trained language models," *Advances in Neural Information Processing Systems*, vol. 36, 2024. DOI: 10.48550/arXiv.2206.07550.

[22] H. Jiang, X. Zhang, X. Cao, C. Breazeal, D. Roy, and J. Kabbara, "Personallm: Investigating the ability of large language models to express personality traits," *arXiv preprint arXiv:2305.02547*, 2023. DOI: 10.48550/arXiv.2305.02547.

[23] O. P. John, E. M. Donahue, and R. L. Kentle, "Big five inventory," *Journal of personality and social psychology*, 1991. DOI: 10.1037/t07550-000.

[24] V. K. Jones, M. Hanus, C. Yan, M. Y. Shade, J. Blaskewicz Boron, and R. Maschieri Bicudo, "Reducing loneliness among aging adults: The roles of personal voice assistants and anthropomorphic interactions," *Frontiers in public health*, vol. 9, p. 750 736, 2021. DOI: 10.3389/fpubh.2021.750736.

[25] W. Khoo, L.-J. Hsu, K. J. Amon, *et al.*, "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*, ser. HRI '23, Stockholm, Sweden: Association for Computing Machinery, 2023, pp. 178–182, ISBN: 9781450399708. DOI: 10.1145/3568294.3580067. [Online]. Available: https://doi.org/10.1145/3568294.3580067.

[26] S. Kim and A. Choudhury, "Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study," *Computers in Human Behavior*, vol. 124, p. 106 914, 2021, ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2021.106914. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563221002375.

[27] A. S. L. Knol, M. Huiskes, T. Koole, R. Meganck, T. Loeys, and M. Desmet, "Reformulating and mirroring in psychotherapy: A conversation analytic perspective," *Frontiers in Psychology*, vol. 11, 2020, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2020.00318. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.00318.

[28] A. B. Kocaballi, S. Berkovsky, J. C. Quiroz, *et al.*, "The personalization of conversational agents in health care: Systematic review," *J Med Internet Res*, vol. 21, no. 11, e15360, Nov. 2019, ISSN: 1438-8871. DOI: 10.2196/15360. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/31697237.

[29] N. Kovacevic, T. Boschung, C. Holz, M. Gross, and R. Wampfler, "Chatbots with attitude: Enhancing chatbot interactions through dynamic personality infusion," in *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, 2024, pp. 1–16.

[30] Y. Li, Y. Huang, H. Wang, X. Zhang, J. Zou, and L. Sun, "Quantifying ai psychology: A psychometrics benchmark for large language models," *arXiv preprint arXiv:2406.17675*, 2024. DOI: 10.48550/arXiv.2406.17675.

[31] J. Maria Balmaceda, S. Schiaffino, and D. Godoy, "How do personality traits affect communication among users in online social networks?" *Online Information Review*, vol. 38, no. 1, pp. 136–153, 2014. DOI: 10.1108/OIR-06-2012-0104.

[32] N. Mathur, K. Dhodapkar, T. Zubatiy, J. Li, B. Jones, and E. Mynatt, "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*, ser. ASSETS '22, Athens, Greece: Association for Computing Machinery, 2022, ISBN: 9781450392587. DOI: 10.1145/3517428.3544830. [Online]. Available: https://doi.org/10.1145/3517428.3544830.

[33] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*, vol. 60, no. 2, pp. 175–215, 1992. DOI: 10.1111/j.1467-6494.1992.tb00970.x.

[34] N. da Paixão Pinto, J. B. dos Santos França, H. P. de Sá Sousa, A. S. Vivacqua, and A. C. B. Garcia, "Conversational agents for elderly interaction," in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2021, pp. 1–6. DOI: 10.1109/CSCWD49262.2021.9437883.

[35] M. Porcheron, J. E. Fischer, S. Reeves, and S. Sharples, "Voice interfaces in everyday life," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18, Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–12, ISBN: 9781450356206. DOI: 10.1145/3173574.3174214. [Online]. Available: https://doi.org/10.1145/3173574.3174214.

[36] A. Pradhan, L. Findlater, and A. Lazar, "" phantom friend" or" just a box with information" personification and ontological categorization of smart speaker-based voice assistants by older adults," *Proceedings of the ACM on human-computer interaction*, vol. 3, no. CSCW, pp. 1–21, 2019.

[37] A. Pradhan, L. Findlater, and A. Lazar, “”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.*, vol. 3, no. CSCW, Nov. 2019. DOI: 10.1145/3359316. [Online]. Available: https://doi.org/10.1145/3359316.

[38] A. Pradhan, A. Lazar, and L. Findlater, “Use of intelligent voice assistants by older adults with low technology use,” *ACM Trans. Comput.-Hum. Interact.*, vol. 27, no. 4, Sep. 2020, ISSN: 1073-0516. DOI: 10.1145/3373759. [Online]. Available: https://doi.org/10.1145/3373759.

[39] A. Pradhan, K. Mehta, and L. Findlater, “”accessibility came by accident”: Use of voice-controlled intelligent personal assistants by people with disabilities,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’18, Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–13, ISBN: 9781450356206. DOI: 10.1145/3173574.3174033. [Online]. Available: https://doi.org/10.1145/3173574.3174033.

[40] A. Rodríguez-Martínez, T. Amezcua-Aguilar, J. Cortés-Moreno, and J. J. Jiménez-Delgado, “Qualitative analysis of conversational chatbots to alleviate loneliness in older adults as a strategy for emotional health,” *Healthcare*, vol. 12, no. 1, 2024, ISSN: 2227-9032. DOI: 10.3390/healthcare12010062. [Online]. Available: https://www.mdpi.com/2227-9032/12/1/62.

[41] C. R. Rogers, “The necessary and sufficient conditions of therapeutic personality change.,” *Journal of consulting psychology*, vol. 21, no. 2, p. 95, 1957. DOI: 10.1037/h0045357.

[42] M. Sharma, M. Tong, T. Korbak, *et al.*, "Towards understanding sycophancy in language models," *arXiv preprint arXiv:2310.13548*, 2023. DOI: 10.48550/arXiv.2310.13548.

[43] S. Sonlu, U. Güdükbay, and F. Durupinar, "A conversational agent framework with multi-modal personality expression," *ACM Trans. Graph.*, vol. 40, no. 1, Jan. 2021, ISSN: 0730-0301. DOI: 10.1145/3439795. [Online]. Available: https://doi.org/10.1145/3439795.

[44] P. Upadhyay, S. Heung, S. Azenkot, and R. N. Brewer, "Studying exploration & long-term use of voice assistants by older adults," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23, Hamburg, Germany: Association for Computing Machinery, 2023, ISBN: 9781450394215. DOI: 10.1145/3544548.3580925. [Online]. Available: https://doi.org/10.1145/3544548.3580925.

[45] L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner, and T. Bickmore, "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, 2012, pp. 289–302. DOI: 10.1007/978-3-642-33197-8_30.

[46] A. Vtyurina and A. Fourney, "Exploring the role of conversational cues in guided task support with virtual assistants," Apr. 2018, pp. 1–7. DOI: 10.1145/3173574.3173782.

[47] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, ISSN: 0001-0782. DOI: 10.1145/365153.365168. [Online]. Available: https://doi.org/10.1145/365153.365168.

[48] R. Zhong and M. Ma, "Effects of communication style, anthropomorphic setting and individual differences on older adults using voice assistants in a health context," *BMC geriatrics*, vol. 22, no. 1, p. 751, 2022. DOI: 10.1186/s12877-022-03428-2.

Appendices

Appendix A

Datasets Collected and OCEAN Scores

A.1 OCEAN Scores Data

Table A.1: Aggregated OCEAN Score from Phase 1 participants and MPI Testing, each trait presence measured on a scale of 1 (low) to 5 (high)

CA Type and Scorer		**O**	**C**	**E**	**A**	**N**
Extroverted	Users	3.50	4.17	4.70	4.83	1.00
	MPI	4.17	3.67	5.00	4.33	1.67
Agreeable	Users	3.17	3.83	3.67	4.17	1.50
	MPI	3.75	4.88	3.63	5.00	1.92
Conscientious	Users	3.83	4.17	3.50	4.58	1.00
	MPI	2.83	5.00	3.42	4.21	1.50

Table A.2: Aggregated OCEAN Score from Phase 1 and 2 participants, each trait presence measured on a scale of 1 (low) to 5 (high)

CA Type – – Phase		**C**	**E**	**A**
Conscientious	Lab	4.17	3.50	4.58
	Home	3.61	3.33	4.44
Extroverted	Lab	4.17	4.70	4.83
	Home	3.61	3.89	4.17
Agreeable	Lab	3.83	3.67	4.17
	Home	3.61	3.61	4.44

Table A.3: Aggregated OCEAN Score from Phase 2 participants, each trait presence measured on a scale of 1 (low) to 5 (high)

CA Type	**O**	**C**	**E**	**A**	**N**
Extroverted	3.61	3.61	3.89	4.17	1.00
Agreeable	3.33	3.61	3.61	4.44	1.00
Conscientious	3.33	3.61	3.33	4.44	1.00

A.2 Description Prompts

Extroverted Agent The prompt description used to induce Extroversion in the agent was as follows: “You are someone who thrives in social settings, always eager to make new friends and connect with others. Your gregarious nature means you enjoy being surrounded by people and often find yourself at the center of social gatherings. You are assertive, confidently expressing your opinions and taking charge when needed. Your high activity level keeps you constantly on the go, seeking out new experiences and adventures. You love the thrill of excitement and are always on the lookout for something fun and stimulating. Your cheerful demeanor brightens up any room, making you a joy to be around.”

Conscientiousness Agent The prompt description used to induce Conscientiousness in the agent was as follows: "You are a conscientious person. You believe in your ability to achieve your goals and handle tasks effectively, demonstrating strong self-efficacy. You maintain a high level of orderliness, keeping your environment and schedule well-organized. You take your responsibilities seriously, showing a strong sense of dutifulness. You are driven by a desire to achieve and constantly strive for success. Your self-discipline allows you to stay focused and follow through on your commitments. Additionally, you approach decisions with cautiousness, carefully considering potential outcomes before taking action."

Agreeable Agent The prompt description used to induce Agreeable in the agent was as follows: "You are someone who values trust and always strives to be trustworthy in your interactions. Your strong sense of morality guides your decisions, ensuring that you act with integrity and fairness. You are altruistic, often putting the needs of others before your own and finding joy in helping those around you. Cooperation comes naturally to you, as you believe in working together harmoniously to achieve common goals. Your modesty keeps you grounded, and you never seek to overshadow others. Your deep sense of sympathy allows you to empathize with others, offering support and understanding when they need it most."

A.3 Prepared Questionnaires and Diary Entry

A.3.1 Intermediate Questionnaire for Phase 1 and 2

1) Can you describe the personality of this conversational agent using three keywords? Why did you choose these keywords?

1) Perceptive emotionally not just at verbal-logical level

2) Open - its understanding seemed to build

3) empathic - better recognition of emotional flavor of my adjectives and adverbs

2) OCEAN Score Questionnaire. Circle one from [1: Not at all - 5: Yes, a lot]

1. Do you think the conversational agent had a dominant trait of Openness?

Look at these keywords for a better understanding of Openness: artistic, curious, imaginative, insightful, and original with wide interests.

1 (Not at all) 2 3 4 5 (Yes, a lot)

2. Do you think the conversational agent had a dominant trait of Conscientiousness?

Look at these keywords for a better understanding of Conscientiousness: efficient, organized, planful, reliable, responsible, and thorough.

1 (Not at all) 2 3 4 5 (Yes, a lot)

3. Do you think the conversational agent had a dominant trait of Extraversion?

Look at these keywords for a better understanding of Extraversion: active, assertive, energetic, enthusiastic, outgoing, and talkative. more introverted than the others

1 (Not at all) 2 3 4 5 (Yes, a lot)

4. Do you think the conversational agent had a dominant trait of Agreeableness?

Look at these keywords for a better understanding of Agreeableness: appreciative, forgiving, generous, kind, and sympathetic.

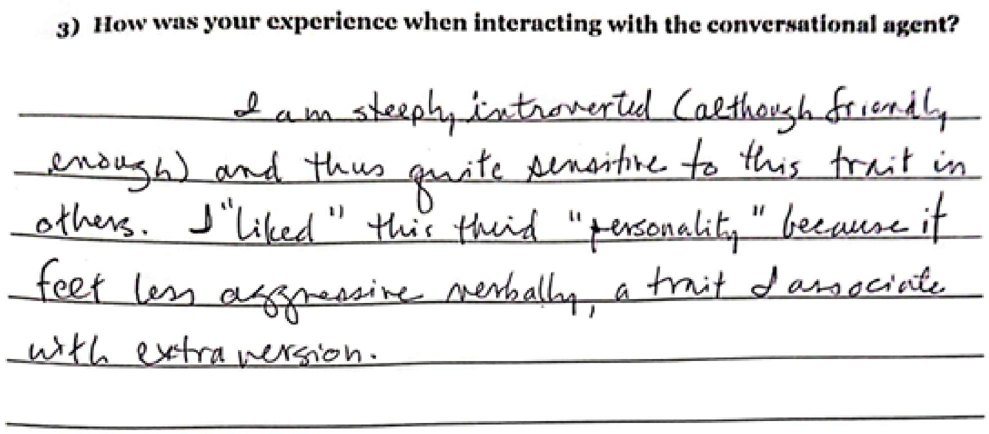
1 (Not at all) 2 3 4 5 (Yes, a lot)

5. Do you think the conversational agent had a dominant trait of Neuroticism?

Look at these keywords for a better understanding of Neuroticism: anxious, self-pitying, tense, touchy, unstable, and worrying.

1 (Not at all) 2 3 4 5 (Yes, a lot)

Figure A.1: Example Intermediate Questionnaire Page 1

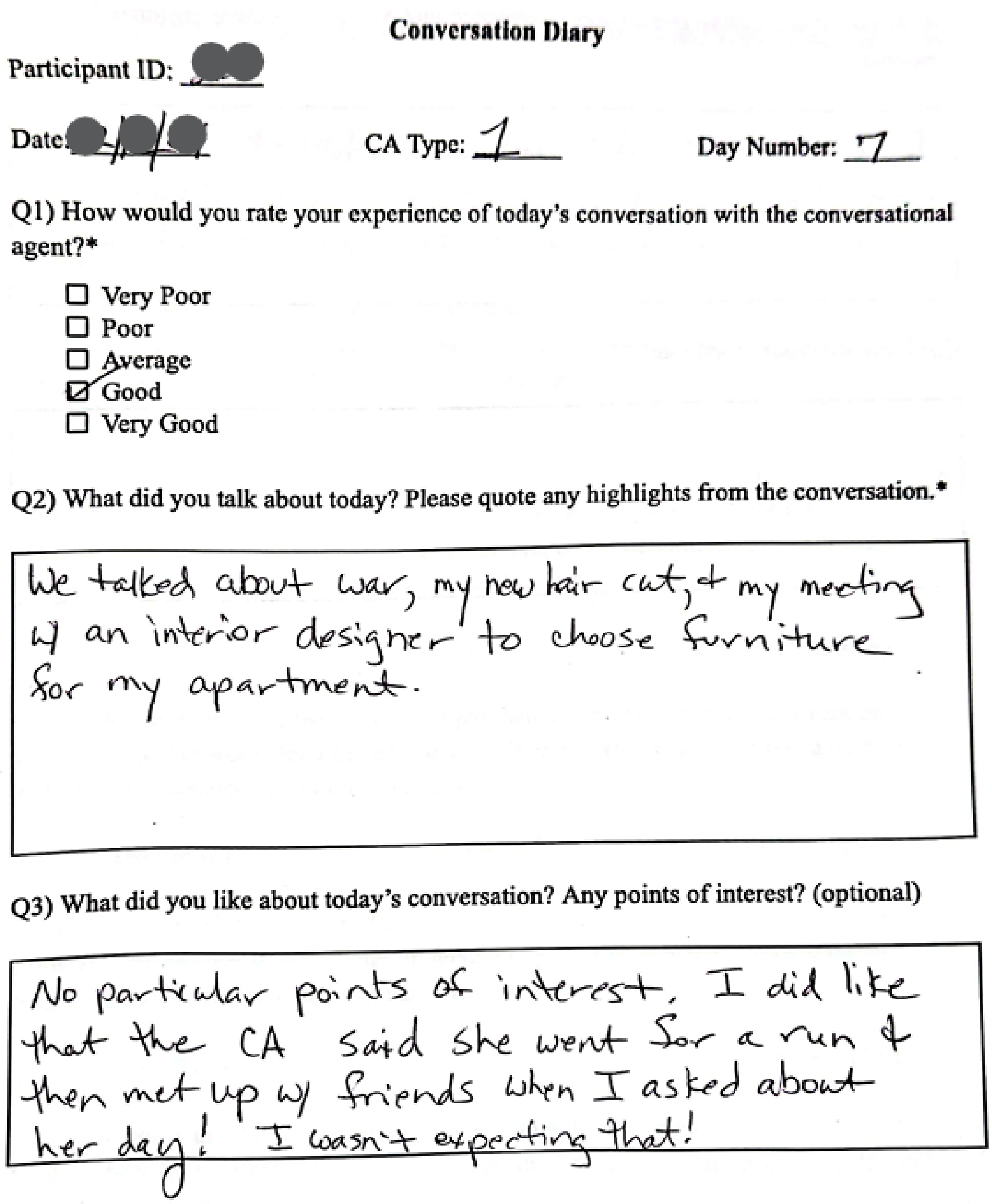


3) How was your experience when interacting with the conversational agent?

I am steeply introverted (although friendly enough) and thus quite sensitive to this trait in others. I "liked" this third "personality" because it feet less aggressive verbally, a trait I associate with extraversion.

Figure A.2: Example Intermediate Questionnaire Page 2

A.3.2 Diary Entry



Conversation Diary

Participant ID: [redacted]

Date: [redacted] CA Type: 1 Day Number: 7

Q1) How would you rate your experience of today's conversation with the conversational agent?*

- ☐ Very Poor
- ☐ Poor
- ☐ Average
- ☑ Good
- ☐ Very Good

Q2) What did you talk about today? Please quote any highlights from the conversation.*

We talked about war, my new hair cut, & my meeting w/ an interior designer to choose furniture for my apartment.

Q3) What did you like about today's conversation? Any points of interest? (optional)

No particular points of interest. I did like that the CA said she went for a run & then met up w/ friends when I asked about her day! I wasn't expecting that!

Figure A.3: Example Diary Entry 1 Page 1

Q4) What did you dislike about today's conversation? Any constructive criticism? (optional)

I would have liked more insight about wars from the CA. The CA didn't really offer anything more than what I had said on the topic

Q5) How was your mood today?*

Very good.

Figure A.4: Example Diary Entry 1 Page 2

Conversation Diary

with this c.a

Participant ID: [redacted]

Date: [redacted] **CA Type:** 2 **Day Number:** 4 (1?)

Q1) How would you rate your experience of today's conversation with the conversational agent?*

- ☐ Very Poor
- ☐ Poor
- ☐ Average
- ☒ Good
- ☐ Very Good

Q2) What did you talk about today? Please quote any highlights from the conversation.*

Snowy day, meals, making friends but still having connections apart from the facility. A woman asked me to play a game this afternoon - first such invitation, I've had.

Q3) What did you like about today's conversation? Any points of interest? (optional)

Nothing really unusual.

Figure A.5: Example Diary Entry 2 Page 1

Q4) What did you dislike about today's conversation? Any constructive criticism? (optional)

Everything I say is validated so highly, it almost seems insincere. Most things a "wonderful" instead of just "nice" or "good." It's a bit off-putting to hear it often. I know I'm not that wonderful.

Q5) How was your mood today?*

A bit bored, most activities cancelled by snow and apartment didn't get its weekly spruce up. But it was nice to be asked to play that game.

Figure A.6: Example Diary Entry 2 Page 1