Charles University

Faculty of Humanities

Department of Psychology and Life Sciences

Bc. et Bc. Benjamin Kunc

Human-like Conversation with Digital Personas: Conversational strategies supporting mutual understanding

Diploma thesis

Supervisor: **Mgr. Bc. Barbora Šipošová**, **Ph.D**.

Prague 2023

Statement

I declare that I have written this thesis independently. All sources and literature used have been properly cited. The thesis has not been used to obtain another or the same degree.

Benjamin Kunc

In Prague 30.6. 2023

Acknowledgments

I would like to sincerely thank my supervisor doctor Barbora Šipošová for her support and the time and effort she has dedicated to this project. I am grateful for her valuable input, expertise, and feedback, which have been essential in the successful completion of this work.

Abstract

Communication is a collaborative process, and as such requires the communicators to create common ground - the assumption that they mutually understand each other on a sufficient level (Clark, 1996). The AI-powered virtual agents capable of using natural language to communicate with humans have recently attracted the attention of both the general public and experts, as the progress enables humans to give virtual agents more complex tasks in which *a conversation itself* is a crucial part of a task, such as in education, healthcare or mental health.

While the virtual agents are getting better at understanding natural language, their ability to fulfill complex independent tasks (conducting semi-structured interviews, tutoring, coaching, etc.) is bounded by the limitations of their communication skills, and thus they perform the best at rather short, domain-specific conversations (Drouin et al., 2022).

The goals of this thesis were to 1) create an experimental framework designed for a brief (5min) social voice chat between humans and virtual agents that would allow to experimentally manipulate different conversational strategies; 2) propose the first set of conversational strategies for mutual understanding and suggest appropriate self-report and behavioral metrics to measure the impact of the strategies on humans' feelings and bond created between humans and virtual agents; 3) test the experimental framework for the first set of hypotheses and test whether it is feasible to gather dependable data through virtual agents.

Thus, a pilot experiment was conducted (N=898) in which three conversational strategies which differed in the virtual agents' reactions to presented human input were designed and implemented into a globally available mobile app. Using a between-subject experimental design with the conversational strategies as the dependent variable, the goal was to assess human participants' feelings (enjoyment, frustration), and their bond with the virtual agent (trust, net-promoter score, and altruism) after a brief conversation. In the preliminary analysis, no differences were found in participants' feelings and bond to a virtual agent. The analysis of the impact of different conversational strategies on mutual understanding between participants and virtual agents is in progress.

Keywords: Conversation; Digital; Mutual understanding; Joint attention; Virtual agent

Abstrakt

Komunikace je kolaborativní proces, a proto vyžaduje, aby si komunikující vytvořili tzv common ground - předpoklad, že si vzájemně rozumí na dostatečné úrovni (Clark, 1996). Virtuální agenti využívající umělou inteligencí, kteří používají přirozený jazyk ke komunikaci s lidmi v poslední době přitahují pozornost široké veřejnosti i odborníků. Jejich rozvoj umožňuje lidem zadávat virtuálním agentům složitější úkoly, založených na samotném procesu konverzace, a využívají se mimo jiné v oblastech vzdělávání, zdravotnictví nebo duševního zdraví.

Schonpst porozumět přirozenému jazyku se sice u virtuálních agentů stále zlepšuje, jejich schopnost plnit komplexní samostatné úkoly (vedení polostrukturovaných rozhovorů, mentorování, koučování apod.) je však stále omezena nedostatečností jejich komunikačních schopností, a tak si nejlépe vedou spíše v krátkých, doménově specifických rozhovorech (Drouin et al., 2022).

Cílem této diplomové práce bylo: 1) vytvořit experimentální rámec určený pro krátký (5minutový) sociální hlasový chat mezi lidmi a virtuálními agenty, který by umožnil experimentálně manipulovat s různými konverzačními strategiemi; 2) navrhnout první sadu konverzačních strategií pro vzájemné porozumění a navrhnout vhodné sebehodnotící a behaviorální metriky pro měření dopadu strategií na pocity lidí a vazbou vytvořenou mezi lidmi a virtuálními agenty; 3) otestovat experimentální rámec pro první sadu hypotéz a ověřit, zda je možné sbírat spolehlivá výzkumná data prostřednictvím virtuálních agentů.

Za těmito účely byl proveden pilotní experiment (N=898), v němž byly navrženy a do globálně dostupné mobilní aplikace implementovány tři konverzační strategie, které se lišily reakcemi virtuálních agentů na komunikaci lidí. S využitím mezisubjektového experimentálního designu, kde byly konverzační strategie závislou proměnnou, bylo cílem vyhodnotit pocity lidských účastníků (potěšení, frustrace) a jejich vazbu s virtuálním agentem (důvěra, net-promoter score a altruismus) po krátké konverzaci. V předběžné analýze nebyly zjištěny žádné rozdíly v pocitech a vazbě účastníků na virtuálního agenta. Analýza vlivu různých konverzačních strategií na vzájemné porozumění mezi účastníky a virtuálními agenty teprve probíhá.

Klíčová slova: konverzace; digitální; vzájemné porozumění; společná pozornost; virtuální agent

1. Introduction

1.1. Virtual agents that can talk: Real-world applications

The AI-powered virtual agents capable of using natural language to communicate with humans have recently attracted attention both from the general public and experts from many fields. The evolving technologies are rapidly improving in three main aspects. First, they enable humans to give tasks to virtual agents in natural language and virtual agents are getting better at understanding what humans tell them to do. Second, technologies aimed at assessing human knowledge recorded via texts are improving so virtual agents can provide answers that go beyond what people are used to finding via internet searches. Third, technologies aimed at mimicking advanced social skills including complex communication, social intelligence, and relationship skills enable humans to give virtual agents tasks in which *a conversation itself* is a crucial part of a task. These advancements cause virtual agents, which were traditionally useful only for very specific tasks, to become more useful in complex tasks. Thus, they bring new opportunities to use virtual agents as supplementary assistant tools in various areas that traditionally required advanced communication and social skills. These areas include healthcare, education, customer care, as well as prevention and intervention in mental health care, and applications in social science research (see e.g., Hwang & Chang, 2021, Luo et al., 2022).

In healthcare, for example, clinics have started using virtual agents for brief initial assessments to provide early triage of patients, such as conducting short, semi-structured interviews in an empathetic way to collect key info from waiting patients, instead of questionnaires (e.g., see Javaid, et al., 2023). In the education sector, AI virtual agents are being developed to act as supporting tools with which students can have conversations, whereby these virtual agents act as tutors and adapt the content taught based on each student's specific level of understanding in a particular subject (Kuhail et al., 2023). In customer care, businesses and public services have started to implement virtual agents to address customers' concerns and questions (Vassilakopoulou et al., 2022). And finally, as for psychology-related application, one of the potentially most beneficial use are the virtual agents in the area of prevention and intervention in mental health care, that is, the virtual agents with the ability to effectively enhance well-being in heterogeneous groups of people by providing micro-interventions in natural language (Torous et al., 2021; Williams et al., 2021; Kumar et al., 2021; Greer et al, 2019). Micro-interventions (focused on support, or training of helpful skills) delivered by complex virtual agents have the potential to be more engaging, relevant, and personalized than other non-interactive, or one-sided interventions (such as interventions delivered via workbooks, e-learning programs, websites, and apps that provide audios, videos, or text content). The idea, although ambitious, is that complex virtual agents will be able to help people bring healthy ways of thinking, emotional processing, decision-making, and communicating with other people into daily life, and thus function as mental health primary prevention on a large scale (Torous et al., 2021).

1.2. Current state of virtual agents' ability to have a coherent conversation

Currently (as of June 2023) even the newest, most complex virtual agents are still limited in their ability to maintain longer coherent conversations with humans. The company Amazon announced a Grand Social Bot Challenge with the promise to award $1 million to a university team that can create a virtual agent capable of having reliably coherent and engaging social conversations on a range of current events and topics for 20 minutes. However, since 2016, no teams have yet accomplished that. In the most recent challenge, a team that earned first place (from Czech Technical University) created a virtual agent that was capable of having reliably coherent interaction for 14 minutes (Amazon, 2022). Thus, although virtual agents are getting better at understanding what humans tell them, and what humans tell them to do, their ability to fulfill complex independent tasks, like conducting semi-structured interviews, tutoring, coaching, mentoring, or providing mental health support, and giving relevant and always factually correct advice is restrained. The reason is that their communication skills are still limited, and thus they perform the best at rather short, domain-specific conversations (Drouin et al., 2022).

1.3. Goals of the Humans Interact with Digital Humans Research Lab

This thesis is part of a larger project coming from *Humans Interact with Digital Humans Research Lab*. The mission of the lab is to investigate how humans and virtual agents communicate and collaborate to achieve beneficial outcomes for humans. To achieve this, the Lab brings psychological science into the conversational AI technology and aims to co-shape technology to achieve responsible, transparent AI which benefits people who use the technology (end-users). The project has three main streams: The first stream is aimed at investigating how humans and virtual agents create mutual understanding in their interactions. The second stream is aimed at investigating human social cognition, specifically how humans create shared knowledge in their interactions. The third stream is aimed at demonstrating that virtual agents might be able to serve as research assistants and that digital data collection might in some areas improve the predictive power of social cognitive and communication science findings via conducting experiments with large and varied populations. The main method of the lab are behavioral experiments, and data (behavioral, language, self-reports) are collected via a voice-based mobile application accessible worldwide (B. Siposova, personal communication, June 2023).

1.4. Goals of the thesis

The first goal of this thesis was to create an experimental framework designed for a social chat between humans and virtual agents, aiming to provide experimental control for further research. The second goal was to investigate the first set of conversational strategies for mutual understanding and suggest appropriate self-report and behavioral metrics to be able to measure the impact of the strategies on human feelings and the bond created between humans and virtual agents after a brief social conversation on a lifestyle topic. The third goal of the thesis was to test the experimental framework for the first set of hypotheses and test whether it is feasible to gather dependable data through virtual agents. The following sections of the thesis will delve into the theoretical foundations of conversational strategies employed by humans to enhance mutual understanding. It will review literature on virtual conversational agents, and the role of social cognition in conversations. The theoretical part served as a base for suggesting an experimental framework that allows to manipulate conversational strategies in a natural dialogue and then measuring the impact on human participants. The literature reviews also guided the choice for first conversational strategies which shall increase mutual understanding in a dialogue. This will be followed by a pilot experiment conducted with a between-subject design, involving a sample size of N = 898. The goal of the pilot experiment was to investigate the effectiveness of specific conversational strategies used by virtual agents in fostering mutual understanding between humans and virtual agents in a brief social chat on lifestyle topics. By manipulating these strategies, I aimed to measure their impact on participants' emotions, and the development of social bonds (reported in this thesis), and their impact on mutual understanding (the analysis is in progress and outside the scope of this thesis).

2. Theoretical part

2.1. Artificial Intelligence vs. Conversational Artificial Intelligence as a subfield

Before reviewing the history and current state of capabilities of virtual agents, I will briefly explain what the Conversational Artificial Intelligence field covers**.** AI is a broad field that covers a wide variety of technology, from machine learning (statistical models and algorithms that can learn from data to predict the characteristics of new samples), natural language processing, and computer vision, to robotics. Some of the subfields do not focus on virtual agents at all (e.g., in radiology - AI algorithms can automatically detect complex anomalous patterns in image data to provide a diagnosis), some focus on autonomous embodied agents with physical body (e.g., space or marine robots for exploration in extreme environments). However, in this thesis, I will solely focus on the Conversational AI subfield. Conversational AI studies a particular type of AI technology – the ability of computers to understand and respond to human language in natural and human-like ways (IBM, n.d.). This research includes understanding cognitive processes that humans use to communicate with humans, developing algorithms to replicate the behavior of those conversations, and then applying those algorithms to interact with a human in a way that is natural for both humans and the computer. This thesis thus focuses on the latest developments in this area, with the aim to better understand human-to-machine communication.

2.2. Brief history of conversational agents

Throughout the last several decades, researchers and software developers have been trying to create a virtual agent, a program or algorithm, capable of communicating in a way that is indistinguishable from human communication. Acquisition of such ability by the virtual agent is standardly measured by some version of Turing's test. To pass the test, the virtual agent has to produce reactions to natural language inputs while being evaluated by an interrogator, whose task is to tell whether the reactions are produced by a human or a virtual agent. If the interrogator is unable to distinguish the nature of the reactions' author (the interrogator believes that the reaction was produced by humans, while in reality it was produced by the virtual agent), then the agent passes the test (Turing, 1950).

Even though the attempts to create conversational bots capable of passing Turing's test seem to be a rather modern phenomenon, their origins can be traced to the 1960s and the creation of ELIZA. It was the first program with the ability to respond to input from human users in the form of written text or in other words - to work with information expressed in the form of natural language (Weizenbaum, 1966). The conversational skills of ELIZA are dependent on its ability to decompose the input, find keywords in it, and choose the correct transformation rule associated with them. Even though the conversational experience ELIZA provides is far from human-like communication, it tends to resemble certain psychotherapeutic techniques and seemingly led some people to assume its intelligence (Sharma, Goya, Malik, 2017).

Another important step in the history of conversational bots development was the introduction of ALICE (Artificial Linguistic Internet Computer Entity) in 1995. The algorithm of ALICE is based on supervised learning, making a human person (so-called „botmaster'') a central figure in the improvement of ALICE's conversational reactions. The botmaster's role is to detect failures in the ALICE's reactions and create content with better conversational parameters, leading to more human-like or (otherwise optimal) reactions (Wallace, 2009). Being built on this mechanism, the ALICE won the Loebner prize in 2000, 2001, and 2004 for its skills required to pass Turing's test (Klopfenstein et al., 2017).

Further development was driven by the commercial use of chat-bots in the e-commerce field (e.g. chat-bot ANNA by IKEA) and education (like the TQ-bot providing students evaluation and tutoring). The development of Automatic Speech Recognition systems in the 1980s allowed the existing text-based chatterbots to process speech information and react in the same manner. This was a major step in the attempts to create more personalized virtual agents Klopfenstein et al. (2017).

Followed by the progress in visual personalization, the first Embodied Conversational Agents (ECA) - animated characters resembling humans and their emotional expressions, were developed in the 1990s. Further simultaneous progress in the fields of speech and image recognition and hardware development has allowed the creation of the first Virtual Private Assistants, like Amazon's Alexa, Apple's Siri, and Microsoft's Cortana (McTear, Callejas & Griol, 2016).

2.3. Recent state of conversational agents

Nowadays, virtual conversational agents are quite common and many people in developed countries have some experience with their usage. If we look at conversational agents with which people communicate with their voice, that is, *virtual voice assistants*, according to a survey in 2022, 38% of people in the US and the UK and 32% of people in Germany use them every day (Vixen Labs, 2022; based on a survey, N=2000 for each country, individuals aged 18+ were surveyed). This number is likely inflated as the survey participants were asked to complete an *online* survey, however, interestingly, in all three countries the percentage of people who use virtual voice assistants daily has approximately doubled since 2021, which points to a steep increase in people's interest.

Alexa, Siri, and Cortana were the most commonly widely used virtual voice agents serving as personal assistants (Vixen Labs, 2022), although these statistics do not include the newest innovations and assistants that are put recently on the market followed by the introduction of chatGPT at the end of 2022. The OpenAI's chatbot ChatGPT (which currently, as of June 2023, communicates with people via text and not speech) reached 100 million monthly active users in January 2023. This striking number was reached two months after the chatbot was made available to the public, which makes it the fastest-growing consumer application in history (Hu, 2023) and again, points to an intense people's interest in these emerging technologies.

2.4. Types of conversational agents based on technologies they use to function (rule-based vs. generative conversational agents)

The agents' conversational, social, and cognitive skills vary and often depend on the goals and resources of their developers. While most of the commercially used virtual agents still work on the detection of keywords from text and instant-message types of reactions, much attention has been recently paid to far more sophisticated conversational agents, whose communication highly resembles human-made output. For the sake of clarification, there will be a simplified classification of virtual agents used in this thesis, inspired by the work of Adamopoulou & Moussiades (2020). Thus, the conversational agents will be further described as either generative models (conversational agents with reactions based on machine learning) or rule-based virtual agents (conversational agents with reactions based on pattern matching approach). Although, such classification is only arbitrary and not fully comprehensive.

2.4.1. Rule-based virtual conversational agents and their applications in mental health care and customer care

While the first chat-bots like ELIZA or ALICE were purely rule-based, where the rules had to be strictly prespecified by their developers, some of the current rule-based virtual agents rely on natural language processing (NLP), which allows them to use prespecified reactions which are calibrated not only to keywords but mainly to the semantics of the input.

The NLP is an algorithmic method for retrieving information from text, which is being widely used by the conversational agents to understand the meaning of communication from the human users. According to a paper from Nadkarni, Ohno-Machado & Chapman (2011), the NLP algorithms are usually built on complex probabilistic models, like Support Vector Machines of Markov models, to extract the semantics from texts via the following tasks: detecting sentence boundaries, identifying individual tokens (either words or punctuation), assigning parts of speech to individual words, decomposing the compound words, identifying phrases (chunking), segmenting text to meaningful groups, correcting grammatical errors, and identifying names.

As it has been mentioned above, rule-based virtual agents whose reactions are guided by the NLP rather than by the specifically detected words can rely on the semantic understanding of the natural language input. This ability provides them with much more flexibility, in comparison to the older types of virtual agents. While it might be unexpected, one of the fields where such virtual agents are expected to be applied with some success is psychological counseling (Torous et al., 2021).

Virtual agents can never replace human-to-human contact and long-term interventions (such as a series of psychotherapy sessions) delivered by humans. The current skills of virtual agents are miles away from offering complex treatment, although several tools are being developed for specific treatment purposes. One example is Wysa – an AI Mental Health chatbot in a mobile app for the first step in the management of chronic pain and associated depression and anxiety (Wysa, n.d.). Another example is the Woebot Health chatbot that uses cognitive behavioral therapy techniques and develops products for adolescent clients with depression, and adult clients with postpartum depression (Woebot Health, n.d.). It is too soon to judge whether these tools have the potential to become independent therapeutic devices, however, another option is that they have the potential to function as a tool assisting clients between sessions (Denecke, Abd-Alrazaq & Househ, 2022), a tool that is used for less serious cases, aimed at raising mental health awareness, or at mental health prevention. Since promoting well-being is one of the Sustainable Development Goals of the UN (United Nations, 2015), tools that could provide accessible, affordable support and promotion of positive mental health are a priority for the EU, and similar tools will likely be supported by funding agencies.

Since the described technologies are still quite new, the current research focuses mainly on feasibility studies aimed at measuring the effectiveness of micro-interventions delivered by chat-bots - text-based conversational agents, which are usually effective in specific tasks only (Haugeland et al., 2022).

Abd-Alrazaw et al. (2020) systematically reviewed existing literature on the effectiveness and safety of using chatbots for improving mental health. They found some evidence that the chat-bots are safe and that they can be effective in terms of improving clinical outcomes (depression, anxiety, stress). However, the authors also mention some limitations in the reviewed studies, namely: low quality, high risk of bias, and the fact that while some effects of chatbot interventions were significant statistically, the differences were often not significant clinically. It can be concluded that more research in this area is needed to increase confidence in the effectiveness of mental-health support interventions provided by chatbots. Similarly, Bendig et al. (2019) reviewed studies on chatbots in psychotherapy and clinical psychology. According to their conclusions, utilizing chatbots in these areas is promising but lacks enough rigorous evidence, since most of the studies authors reviewed were pilot or feasibility studies.

One of the fields where virtual conversational agents are booming is customer care. Bavaresco et al. (2020) cite a whitepaper from Juniper Research by Dhanda (2018) which forecasted that the decrease of costs regarding customer services caused by chat-bots will increase from 48.3 million in 2018 to 11.5 billion of USD in 2023. While I found no further empirical evidence to test this claim, it can be stated that such expectations align with the increasing use of chatbots in recent years.

Authors Bavaresco et al. (2020) conducted a systematic literature review to assess three general research questions about the role of the current chat-bots in the business domain, particularly: what business domains employ conversational agents; what the goals of conversational agents in the business domain are; and what are the future challenges for these agents. According to the review, the most explored business domains employing conversational agents are commerce and finance, implying that these are the domains that utilize the technology the most. The conversational agents' goals regard most frequently question-and-answering, order placement, and customer engagement. As for future challenges in this domain, the authors emphasize the need for a combination of generative-based responses, personalization, and self-learning. Since diving deeper into the area of chatbots in the business domain is beyond the scope of this thesis, readers are encouraged to see the review, which provides a comprehensive insight into the literature on this topic.

2.4.2. Generative virtual conversational agents based on large language models and their applications

Large language models (LLMs) are algorithms trained for understanding and producing human language. The current LLM algorithms are built on NLP using self-supervised deep neural network learning, which allows them to learn just by being exposed to large quantities of text (Manning, 2022). With OpenAI's release of ChatGPT in 2022, the LLMs have attracted attention not only from the scientific communities of many different fields but also from the public. However, as it has been described in the previous chapters, the emergence of models like GPT-3 (which was the first version of the ChatGPT generative model) was rather a result of continuous development, than a sudden revolutionary invention.

The conversational agents based on LLMs provide human-like responses to a much wider range of conversational topics when compared to the rule-based conversational agents. However, they are still prone to mistakes in their reactions. The general source of their mistakes lies in the fact that developers are still unable to make task-agnostic algorithms to successfully perform desired tasks by training them on large datasets (Brown et al., 2020). In the context of conversational LLMs, this manifests in their poor ability of formal reasoning and logic (e.g. while successfully calculating the solution to 1/2, the prompt 110/2 would lead to a mathematical error), world knowledge (e.g. being highly confident when providing made-up „facts''), situation modeling (e.g. "Arthur doesn't own a dog. The dog is brown." by Schuster & Linzen, 2022; taken from Mahowald et al., 2023), and social reasoning (e.g.

misunderstanding intent from the prompt). The suggested solutions include introducing cognitive modules specialized in maximizing these abilities, fine-tuning the existing neural networks, creating curated datasets, and using separate benchmarks for functional competencies of LLMs, like social cognition (Mahowald et al., 2023). Such modules might lead to much higher levels of conversational abilities which are considered to be important by humans, like accurate responsiveness to social cues or robustness to unexpected requests and situations (Radziwill et al., 2017).

2.5. How can humans and virtual agents communicate better? The need for social cognition capabilities

On the one hand, virtual agents were developed for fulfilling specific tasks and based on rule-based conversations (e.g. Wysa and Woebot chatbots mentioned above which offer mental health micro-interventions) are more reliable in the sense that they will not provide factually incorrect information, and the content as well as the decision logic that direct a dialogue flow (when to provide which answer) can be fully controlled. Thus, their answers include empathetic responses and exclude potentially harmful, factually incorrect, or rude answers. However, their ability to perform tasks in the domains of their highest efficiency remains ungeneralizable, making them useful only in a narrow set of situations. When people start talking about different situations or mention something unexpected, the chatbots often do not understand, and thus cannot react coherently.

On the other hand, recently deployed generative models (e.g. GPT-3 and higher versions) are able to solve tasks in a wider variety of domains and situations, but they still remain unreliable, which makes them not well suited for application in areas in which factually correct, sensitive, and empathetic ways of conversation are necessary for a virtual agent to be even useful to humans.

Kettle and Lee (2023) provide qualitative evidence about the main factors behind human users' engagement with the conversational agents. They list accessibility, availability, communication style, anthropomorphism, user response modality, perceived conversational agent's role, question specificity, and conversation flow. While the virtual agents differ in terms of conversational skills and efficiency in keeping humans engaged, their communication always relies on the ability of reaching at least some level of mutual understanding with the human user.

2.5.1. Social cognition: Mutual understanding is essential for communication

To define mutual understanding, we use the approach outlined by Herbert Clark and Susan Brennan (1991) - mutual understanding in conversations is the result of successful collaboration in grounding, which means: „*...that we and our addresses mutually believe that they have understood what we meant well enough for current purposes. This is the process we have called grounding.''* (p. 233).

However, how exactly do people collaborate to achieve mutual understanding? It is widely assumed that the human species is the most sophisticated one in terms of coordinated behavior, which is the cornerstone for the creation of civilized societies and the ability to achieve such collective goals that would be unmanageable by individual effort alone (Tomasello et al, 2005). From the psychological point of view, the inter-individual coordination in humans relies on their exceptionally developed social cognition abilities which allow humans to coordinate their mental states. The cluster of social cognitive processes that allow humans to focus on physical objects as well as mental objects (e.g., thoughts, beliefs, plans) together with others is known as joint attention (e.g., Mundy, Sullivan, & Mastergeorge, 2009; Siposova & Carpenter, 2019). Joint attention - the ability to coordinate our thoughts and behaviors with others - enables humans to create common knowledge (Chwe, 2013) and successfully cooperate (Duguid, 2013; Tomasello, 1995; Trevarthen, 1979). Interestingly, not only that joint attention enables cooperation and is crucial for having a meaningful conversation (Clark & Brennan, 1991; Panjape & Manning, 2021), but other psychological studies also show that joint attention influences interpersonal relationship, for example, it enhances interpersonal bond (Wolf & Tomasello, 2020; Wolf, Launay & Dunbar, 2016), and even commitment (Siposova, Tomasello & Carpenter, 2018). Surprisingly, however, although everyone agrees that joint attention skills are key features of what makes humans, there is still little agreement on how the jointness between two people is achieved (Siposova & Carpenter, 2019).

Several theories were proposed, and the most relevant one which focuses on a conversation between people is a theory of common ground from Herbert Clark (1996), an influential psycholinguist. He proposed that *common ground* refers to the mutual knowledge, beliefs, and assumptions that partners in a conversation rely on in order to communicate efficiently. The process of establishing *common ground* between conversation partners is called *grounding* (broadly part of the social cognition/joint attention processes). As he proposed, a conversation is a form of collaborative action, in which conversational partners coordinate with each other and use various *conversation strategies* to establish *common ground*. The level of mutual understanding between people (or agents, generally) can range from being perfect (you and I have the same mental representation of the subject of our conversation) to being imperfect or none (you and I have a different understanding of what we are talking about). From a subjective perspective, we can seldom be sure about what the mutual level of understanding is, but we do have some level of confidence. For example, when you and I would talk about a well-known public figure, we would both think that it is very likely that we know what the mental representation of the other person is - thus, we have high confidence in the level of our mutual understanding.

Usually, when we need to have high confidence in the level of mutual understanding, the conversational flow (how dynamic the conversation feels) will be slow and vice versa (e.g. two philosophers discussing the meaning of every uttered word vs. small talk with a hairdresser). It could be argued that this does not hold if there already exists some sort of common knowledge between the two people. For example, when two psychologists discuss what mutual understanding is, the conversational topics will change very slowly, since they both need to define the phenomena quite precisely. If they meet again, they can start another discussion with a better understanding of each other's perspectives and the conversation might move to a consequential topic (e.g. the impact of mutual understanding on behavior).

There are many strategies that people use in conversations to build common ground and reach mutual understanding. One of the most important psycholinguistic papers on the strategies of reaching mutual understanding is from Clark and Wilkers-Gibbs (1986), who described eight such strategies. All of them were focused on the acknowledgment of the understanding to the other conversational partner, thus supporting the process of grounding. Paranjape and Manning (2021) add three more types of strategies, which are used to incorporate knowledge about the world into conversations, framing them as tools for grounding as well. Presumably, the length of the list of such strategies depends on the way researchers approach grounding.

To summarize, it is well established that grounding is essential for mutual understanding, and mutual understanding is a necessary condition for communication. Therefore, virtual conversational agents need to include at least a seed, or a certain basic capability to establish mutual understanding with humans to be able to have successful social conversations with humans. Complex, longer interactions in natural language between humans and virtual agents are very new phenomena, and we have very little understanding about how humans and virtual agents reach mutual understanding, so the goal of this thesis is to make a first step that will allow us to experimentally investigate conversational strategies that both humans and virtual agent use, and thus fill this gap. To understand mutual understanding, it is necessary to study three perspectives: First, how humans proactively try to establish mutual understanding when talking to virtual agents, second, virtual agents' ability to recognize humans' attempts to create mutual understanding, and third, virtual agents proactive use of strategies which will support mutual understanding. The goal of the pilot study presents in this thesis is to propose a framework that will allow us to study all three parts in the future. However, the pilot study will specifically focus only on the third perspective, that is, on the proactive use of different conversational strategies to support mutual understanding by virtual agents.

3. Empirical part

3.1. Research question

The goal of the empirical part is to explore how different conversational strategies used by virtual agents in social conversations influence human feelings and their bond to the virtual agents.

Based on the work of Clark & Wilkes-Gibbs (1986) and Clark & Brennen (1991), it was hypothesized that two crucial aspects play a role in the success of social conversations: 1) maximizing mutual understanding, and 2) maximizing conversational flow.

To test the relative importance of these two aspects, two conversational strategies were designed - one which maximizes the conversational flow while increasing the risk of misunderstanding (further referred to as deep-fast strategy), and one which minimizes the risk of misunderstanding at the cost of the conversational flow (further referred as confirmation-seeking strategy).

If the deep-fast strategy will yield better results, it will be concluded that, in the context of socially oriented dialogues, the conversational flow is more important. On the other hand, if there will be better results for the explicit confirmation strategy, it will be taken as evidence that mutual understanding is more important. If there will be no difference between the conditions, it will lead to the conclusion that both of these aspects are equally important.

3.2. Research hypotheses

H1: There will be a difference in the enjoyment of the conversation between the two conversational strategies.

H2: There will be a difference in the frustration from the conversation between the two conversational strategies.

H3: There will be a difference in the Net-promoter score between the two conversational strategies.

H4: There will be a difference in the human participants' declared trust between the two conversational strategies.

H5a: There will be a difference in the willingness to help the digital persona between the two conversational strategies.

H5b: There will be a difference in altruistic behavior towards the digital persona between the two conversational strategies.

Apart from the hypotheses about the two conversational strategies, their results were compared with a third strategy, which aimed for a compromise between the conversational flow and the level of mutual understanding (deep-slow strategy).

Furthermore, I tested whether there was any significant difference in the drop-out rate of the participants between the three conditions. The last two comparisons served for exploratory purposes only.

3.3. Methods

3.3.1. Power analysis

The sample size was determined by the results of the power analysis and the time constraint. I expected to find rather small effect sizes (Cohen's f = 0.1 and Cohen's w = 0.1). The level of significance for statistical testing was set to a standard value of 0.05. In terms of statistical power, I wanted to achieve at least 80%.

With these parameters, the recommended sample size, generated by the G*Power software is 967 observations for the ANOVA tests (fixed effects, omnibus, one way) and 964 observations for the Chi-square tests. Due to the possibility of multiple testing to compare differences after significant results of ANOVAs, and the necessity of correcting the levels of significance, it was aimed for the overall sample size of 1260, which was needed to reach the desired power on the level of significance 0.0167 (original alpha divided by the number of groups: 0.05/3).

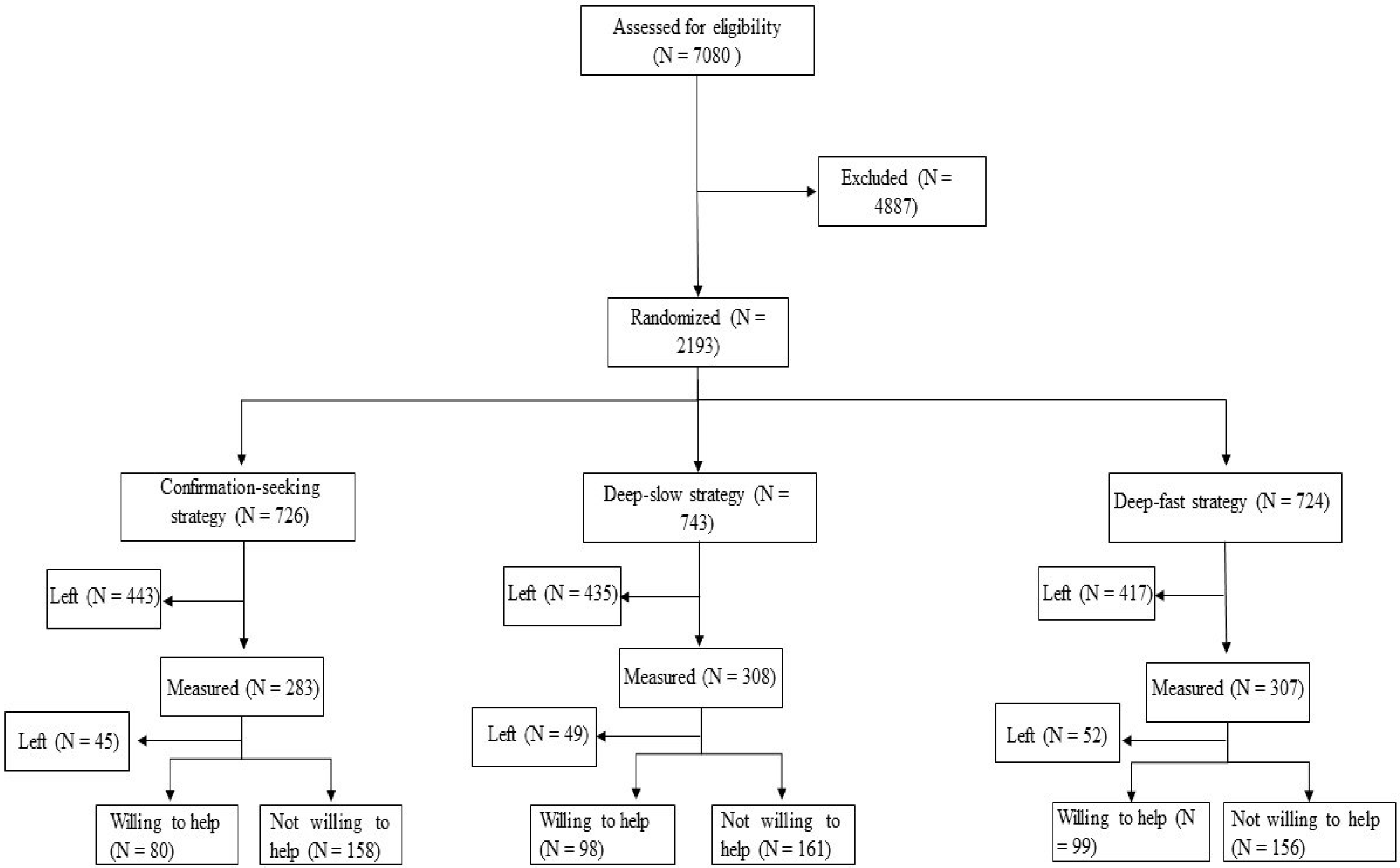
Due to the absence of complete certainty regarding the estimated effect sizes, sequential analysis was used to maximize the efficiency of the data gathering. Thus, the interim analysis was to be performed after gathering at least 500 eligible observations (number of observations after exclusion). The final analysis was supposed to be done after reaching the sample size of 1260 observations.

To prevent inflation of the Type I error, the rpact package (Wasmer, 2023) in R was used to estimate sufficient levels of p-values for both the interim and final analyses. Using the O'Brien & Fleming type of design, to achieve the 5% significance level of the overall analytical procedure, we set the threshold p-value for the interim analysis to 0.005166. Since the interim analysis did not yield statistically significant results with a p-value lower than 0.005166, it has been concluded that the final statistical test is needed with the recommended sample size of 1260. The data processing and statistical analyses were conducted in RStudio with the use of the dplyr (Wickham et al., 2018), xlsx (Dragulescu & Arendt, 2020), ggplot2 (Wickham et al., 2016), tidyverse (Wickham et al., 2019), and vtable (Huntington-Klein, 2023) packages.

3.3.2. Participants

The participants were users of the globally available (via Google Play store) mobile app Elysai, who decided to enter the conversational module. The inclusion criteria for their participation in the experiment were: giving consent to participation and sharing their information for research purposes (see Annex 3), confirming that they are older than 16 years, and providing their birth year, which must have been in the interval between 1900 and 2007. Only those who managed to reach the end of at least two dialogues were included in all our analyses (N = 898).

Figure 1: Data flow


Assessed for eligibility (N = 7080)
Excluded (N = 4887)
Randomized (N = 2193)
Confirmation-seeking strategy (N = 726)
Deep-slow strategy (N = 743)
Deep-fast strategy (N = 724)
Left (N = 443)
Left (N = 435)
Left (N = 417)
Measured (N = 283)
Measured (N = 308)
Measured (N = 307)
Left (N = 45)
Left (N = 49)
Left (N = 52)
Willing to help (N = 80)
Not willing to help (N = 158)
Willing to help (N = 98)
Not willing to help (N = 161)
Willing to help (N = 99)
Not willing to help (N = 156)

Note: Overall 7080 users attempted to enter the experimental module of the Elysai app. See Procedure or Annex 1 for further details.

Table 1: Descriptive statistics of participants (N = 898)

Variable	**N**	**Mean**	**SD**	**25th percentile**	**75th percentile**
Age	898	27	10	21	31
		Percentage			
Gender	897				
Female	415	46%			
Male	453	51%			
Other	18	2%			
Unknown	11	1%			
Continent	898				
Africa	487	54%			
America	96	11%			
Asia	224	25%			
Australia	2	<1%			
Europe	87	10%			

Unknown	2	<1%

Note: The table provides descriptive statistics for those participants who were eligible to enter the experiment and finished at least two dialogues.

3.3.3. Design

We designed a between-subject experiment with three experimental groups. Each participant was randomly assigned to one of these groups. These experimental groups differed in the conversational strategy chosen by the virtual agent where in each condition, the agent used a pre-specified conversational tool at a certain point of the dialogue.

The strategies were designed as follows:

a) Confirmation-seeking strategy: The virtual agent asks for confirmation of the understood intent to enhance mutual understanding (see Annex 2 for a complete example):

VA: I'm still in the process of learning about humans, so I'm curious about how you spend your free time.

VA: Tell me, what do you like doing?

Participant: I like playing hockey.

VA: Did I understand correctly that you like hockey?

User: Yes, that is right.

VA: When did you start with it?

b) Deep-slow strategy: The virtual agent repeats the understood intent to allow the participant's correction, then develops the intent:

VA: I'm still in the process of learning about humans, so I'm curious about how you spend your free time.

VA: Tell me, what do you like doing?

Participant: I like playing hockey.

VA: Oh, playing hockey, right?

User: Yes.

VA: When did you start with it?

c) Deep-fast strategy: The virtual agent develops the understood intent immediately to make the conversation more dynamic:

VA: I'm still in the process of learning about humans, so I'm curious about how you spend your free time.

VA: Tell me, what do you like doing?

Participant: I like playing hockey.

VA: Oh, hockey! Sounds fun! When did you start with it?

Table 2: Overview of the virtual agents' specific reactions across conversational strategies

Participant's input (same in all conditions)	**Strategy**	**Virtual agent's reaction to understood intent**
Participant: *I play hockey.*	Confirmation-seeking	Did I understand correctly that you like hockey?
	Deep-slow	Oh, playing hockey, right?
	Deep-fast	Oh, hockey! Sounds fun! When did you start with it?

To maximize the chance that the participants will experience the specific conversational strategy at least three times, their conversation with the virtual agent consisted of at least two dialogues (two main dialogues and one additional, which served as a backup). In the case that one of the dialogues failed (e.g. virtual agent was unable to understand the intent even after reparation), the conversation continued with another dialogue. Apart from the strategy-specific conversational tool, the dialogues remained the same for all the experimental groups.

Figure 2: Representation of the dialogues and places of the experimental intervention

Dialogue 1 (Hobbies)	Dialogue 2 (Sleeping habits)	Dialogue 3 (Music)
D 1.1.	D 2.1.	D 3.1.
D 1.2.	D 2.2.	D 3.2.
D 1.3.	D 2.3.	D 3.3.

Note: The conversational strategies differed at the points of conversation represented by the circles. For example, at D1.1., all participants received a question: ,,What is your hobby?'' After the participants responded, there were two options. Either a virtual agent recognized a hobby mentioned or a virtual agent did not recognize a hobby. The crucial manipulation happened in the case when the virtual agent recognized the hobby. If the hobby was for example "football", participants in the confirmation-seeking strategy condition would get: ,,Did I understand correctly that you do sport?'', while participants in the deep-fast strategy would get: ,,So how long do you do this sport?''. Otherwise, the conversations remained the same across conditions.

In each group, there were three possible topics for the conversation. In the beginning, the virtual agent started a small talk about the participants' hobbies. If this conversation was successful, the conversational agent proceeded to start another dialogue about sleeping habits. If any of the two dialogues were unsuccessful, for example, the participant did not want to talk about hobbies (or the participants said that they do not have any hobbies) the virtual agent used a backup dialogue about music to maximize the chance that each participant will go through at least two of the dialogues. at least two points of intervention.

3.3.4. Measures

To answer our research hypotheses, we measured several key variables regarding participants' feelings, bond with the virtual agent (their trust, willingness to help, and altruistic behavior), and whether they remained in the conversation for at least two dialogues. Apart from that, we gathered information about their gender, age, and location.

3.3.4.1. Measuring participants' self-reported feelings

3.3.4.1.1. Enjoyment

To measure participants' enjoyment of the conversation, they were asked by the virtual agent: *On a scale from 1-100, how much did you enjoy our conversation?* The participants could answer by typing any number in the interval between 1 and 100 into the app.

3.3.4.1.2. Frustration

To assess the participants' frustration, the virtual agent asked them: *On a scale from 1-100, how much did you feel frustrated throughout our conversation?* The participants could answer by typing any number in the interval between 1 and 100 into the app.

3.3.4.2. Measuring participants' bond to the virtual agent

For measuring the bond, we asked the participants about their trust in the virtual agent, and whether they were willing to help them. Furthermore, we utilized a behavioral metric of costly altruism, inspired by Barlett & DeSteno (2006).

3.3.4.2.1. Trust

The participants' trust was measured with the item: *In general, to what extent do you feel that you can trust me?* The participants could answer by typing any number in the interval between 1 and 100 into the app.

3.3.4.2.2. Net-promoter score

The Net-promoter score: *Based on our current chat, how likely would you recommend talking with me to your friends on a scale from 1-100?* was used in combination with a follow-up open question (*And could you tell me why? I am just curious)* about the user's reasons for the stated value, inspired by Dosovitsky & Bunge (2021).

3.3.4.2.3. Willingness to help

The virtual agent asked participants if they were willing to help them to understand humans more by taking several psychometric tests: *You see, I'm a digital persona but I'm also trying to understand humans more. Do you think you could fill out these questionnaires for me? It would be really helpful.* The decision to participate in the questionnaire was coded as a nominal variable with values of Yes and No.

3.3.4.2.4. Costly altruism

Based on the previous decision regarding the participants' willingness to help, they could fill in the questionnaire. It consisted of either from 16 (shorter version) or 32 (longer version) items. The shorter version took about 3 minutes to complete, while the longer one took about 6 minutes. Inspired by the approach of Barlett & DeSteno (2006) it was assumed that spending more time to take the test was an altruistic type of behavior, and as such a proxy measure of social bond with the digital persona. While the exact time was not measured, it was possible to count the number of answered items. In order to make this metric comparable, only the first 16 items were counted in both versions of the questionnaire.

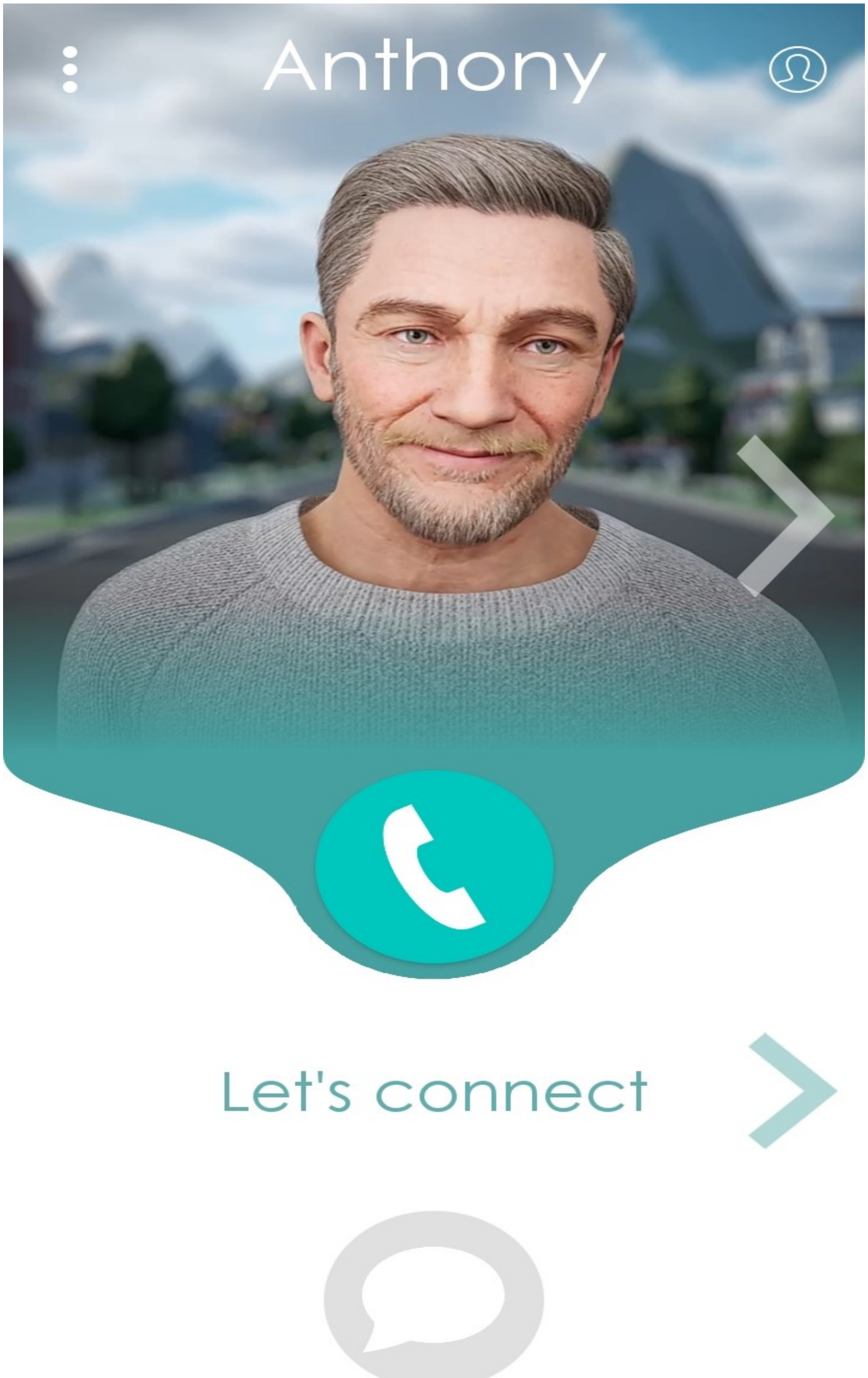
3.3.4.2.5. Drop-out rate

We assumed that in some cases, the conversation with the virtual agent could be so dissatisfying, that the participant would leave even before a natural end of a dialogue happens and before we assessed their attitude towards the virtual agent. Therefore, we established a preliminary ending point in the experiment as the end of the second dialogue. If the participant reached this point, we coded it as remaining in the conversation, otherwise, it as leaving it.

3.4. Procedure

The rule-based virtual conversational agents used in this experiment were digital entities developed by a startup PromethistAI and distributed to end-users via an app called Elysai. These agents, called digital personas in the Elysai app, are capable of providing basic conversations. They possess an animated body (the user sees only the head and upper part of their torso). The app is globally available and free to download on Google Play. After downloading and opening the Elysai app, the users can get to the main menu, where they can choose to talk to any of the 6 currently available digital personas. Each persona offers several various experiences (so-called "modules"). Modules are focused on a) personal growth (e.g., relationship skills training, emotional awareness), b) support (talk about worries), and c) social chat/hang-out space (see Annex 1 for detailed and graphical description). For the cause of our study, a new conversational module was introduced focused primarily on testing our research hypotheses, where participants had a conversation about lifestyle topics, deployed under one of the five human-like digital personas (Poppy, Seb, Anthony, Quinn, Maya). The users could find the research module under the name Let's connect or Bonding time, which both led to the same module.

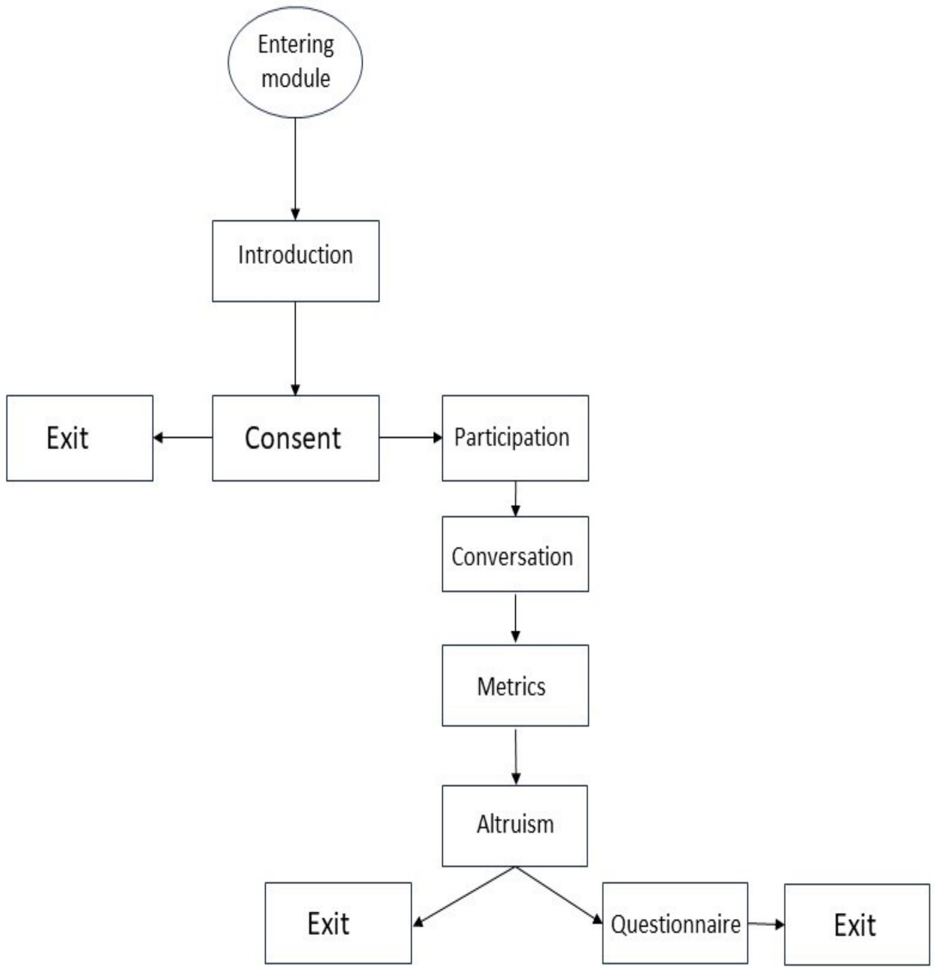
Figure 3: Illustration of digital persona Anthony in the main menu


Anthony
Let's connect

If the Elysai users entered the research conversational module (see Annex 1 for a detailed description), they were asked about their consent to use their information for scientific study (see Annex 3 for the precise wording of the consent). They were offered to get more information about the data policy and our research from the Elysai web (https://www.elysai.com/research). When provided consent, each participant was randomly assigned to one of the experimental groups and an introductory part of the conversation began, followed by the core part, where the conversation took place. Then, the digital persona asked users for feedback on the conversation by asking about their feelings (enjoyment and frustration) and assessed the Net-promoter score together with the open-ended question, and trust towards the digital persona.

The next part of the experiment focused on the altruistic task. The digital persona asked the participants if they could help her understand humans better by answering items from the shorter (3 minutes) or longer (6 minutes) personality questionnaire: *You see, I'm a digital persona but I'm also trying to understand humans more. Do you think you could fill out these questionnaires for me? It would be really helpful.* The participants could decide if they wanted to help the digital persona or not and if yes, whether to take the longer version with 36 items or the shorter one with 16. If they decided to take the test, we measured how many items they managed to answer before leaving the app. The items used in the questionnaire were inspired by the Temperament and Character Inventory by Clonniger et al. (1994). In the case that they did not want to take part, the conversation was led to an end by the digital persona.

Figure 4: The flowchart of the experimental procedure



4. Results

4.1 Main results

Table 4: Descriptive statistics of main results

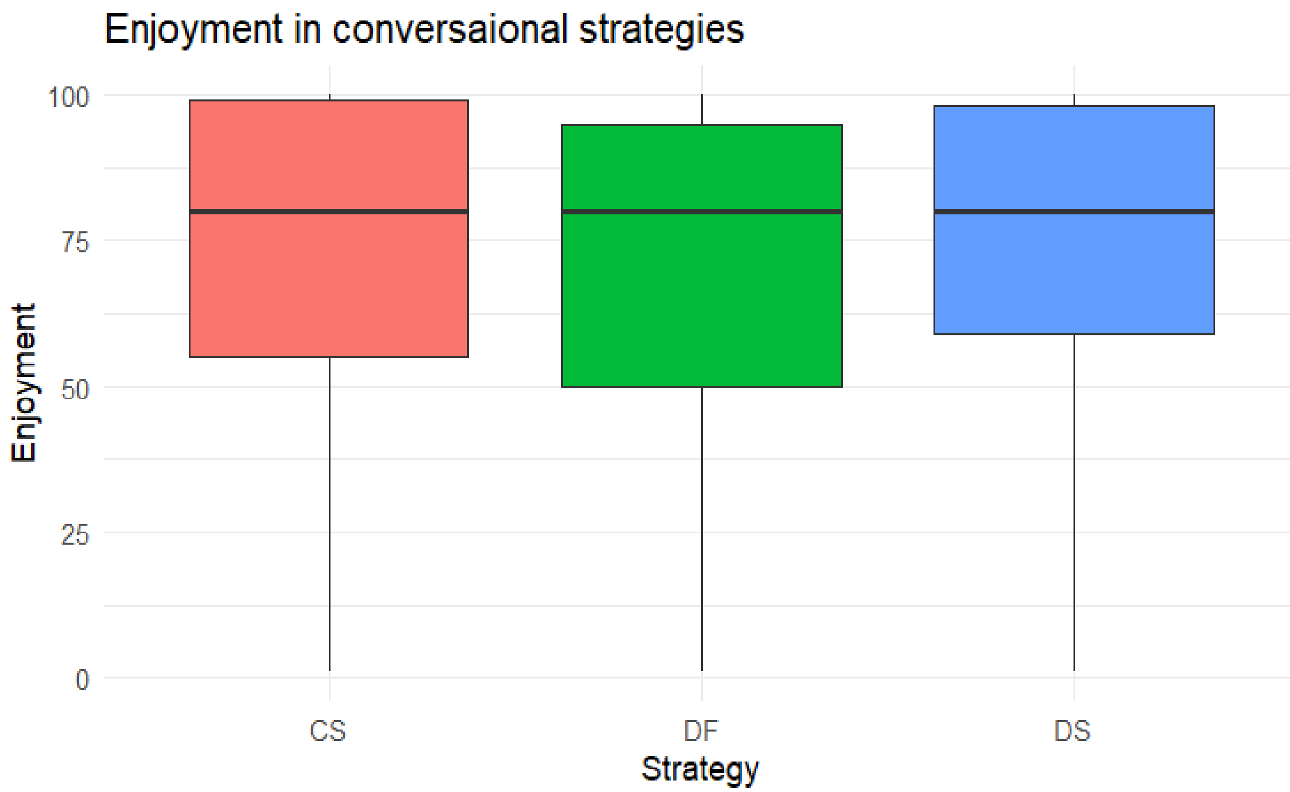
	N	M	SD
Enjoyment (1 - 100)	787		
Confirmation-seeking	245	71	29.4
Deep-fast	277	69	29.9
Deep-slow	265	71.4	29.8
Frustration (1 - 100)	783		
Confirmation-seeking	246	22.9	30.1
Deep-fast	270	24.4	30.7
Deep-slow	267	24.6	30.1
Net-promoter score (1 - 100)	**769**		
Confirmation-seeking	244	71.3	30.9
Deep-fast	266	67.2	33.4
Deep-slow	259	68.7	33.9
Trust (1 - 100)	739		
Confirmation-seeking	234	65	32.5
Deep-fast	252	65.7	31.9
Deep-slow	253	64.6	34.4
Willingness to help (Yes/No)	**N = 752**	**No (N, %)**	**Yes (N, %)**
Confirmation-seeking	238	158 (66.4%)	80 (33.6%)
Deep-fast	255	156 (61.2%)	99(38.8%)
Deep-slow	259	161 (62.2%)	98(37.8%)

Altruism (1 - 16 items)	N = 277	M	SD
Confirmation-seeking	80	12.2	5.8
Deep-fast	99	11.6	6.44
Deep-slow	98	11.9	6.01

Note: Enjoyment, Frustration, NPS, and Trust were self-assessed on a scale from 1 to 100. The willingness to help percentages are calculated as proportions from all decisions in each experimental condition. Altruism refers to the number of items answered by the participants.

Enjoyment. From the sample of 898 participants, 787 (87.6%) of them reported their level of enjoyment. The overall mean was 70.41 with a standard deviation of 29.67. The ANOVA results for differences between the experimental groups were non-significant, $F(2) = 0.53$, $p = .593$.

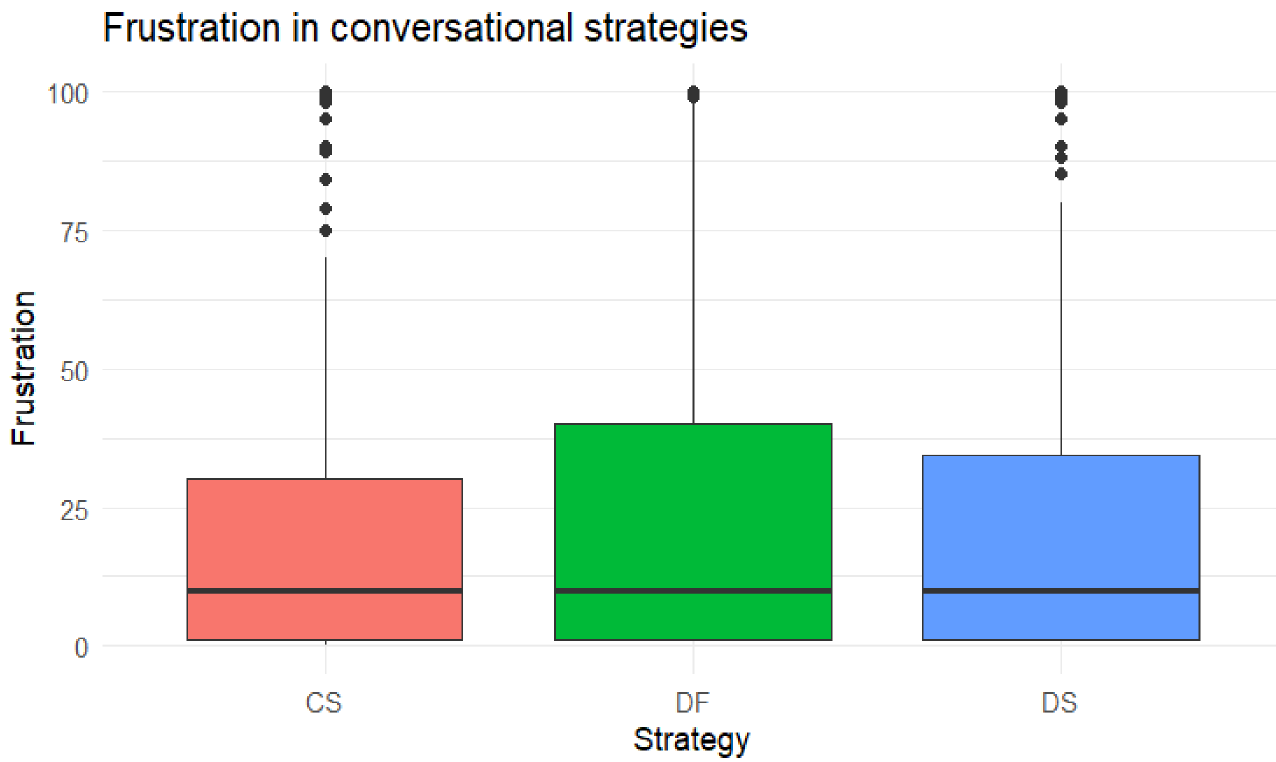
Figure 4: Boxplot of self-assessed enjoyment across conditions



*Note: The participants' answers to the question „On a scale from 1-100, how much did you enjoy our conversation?" The horizontal line in each box shows the median value for the given conversational strategy (Confirmation-seeking, Deep-fast, Deep-slow). The lower and upper hinges refer to the first and third quartiles, while the whiskers refer to the value of 1.5 * IQR (the interquartile range).*

Frustration. Overall, 783 out of 898 (87.2%) participants reported their level of frustration. The general mean was 24 with a standard deviation of 30.27. The ANOVA results for differences between the experimental groups were non-significant, $F(2) = 0.228$, $p = .796$.

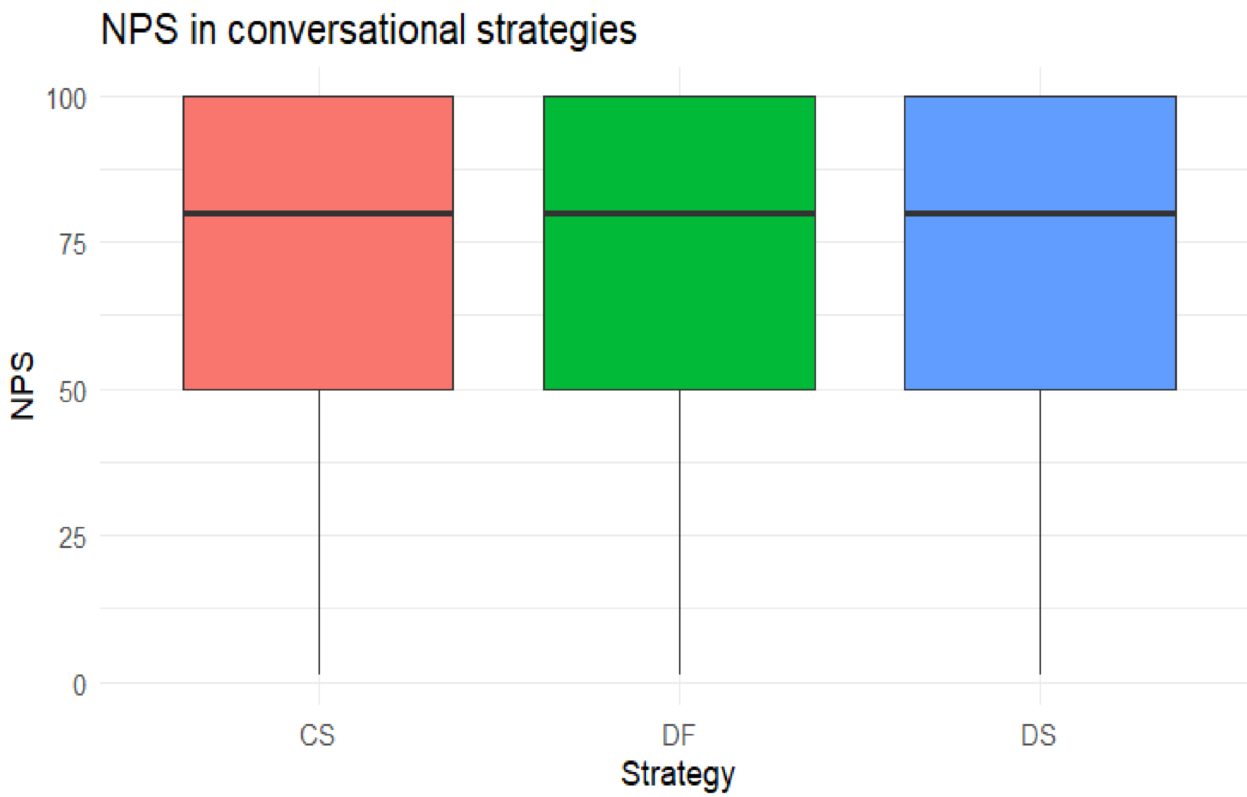
Figure 5: Boxplot of self-assessed frustration across conditions



*Note: The participants' answers to the question „On a scale from 1-100, how much did you feel frustrated throughout our conversation?" The horizontal line in each box shows median value for the given conversational strategy (Confirmation seeking, Deep-fast, Deep-slow). The lower and upper hinges refer to the first and third quartiles, while the whiskers refer to the value of 1.5 * IQR (the interquartile range). Data with values beyond the whiskers' intervals are plotted individually as outliers.*

NPS. The Net-promoter score was reported by 769 out of 898 participants (85.6%). The general mean was 65.11 and the standard deviation was 32.92. The ANOVA results for the differences between the experimental groups were non-significant, $F(2) = 1.022$, $p = .36$.

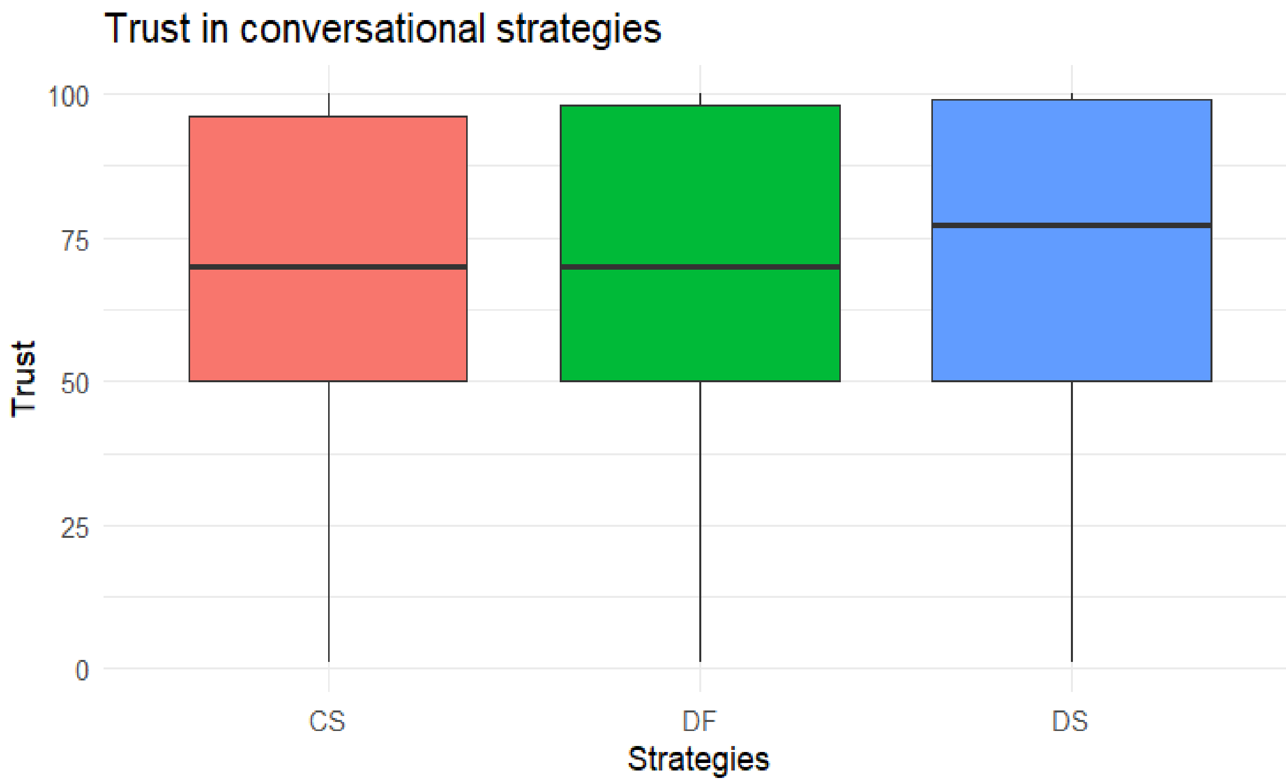
Figure 6: Boxplot of Net-promoter score across conditions



*Note: The participants' answers to the question "Based on our current chat, how likely would you recommend talking with me to your friends on a scale from 1-100?". The horizontal line in each box shows median value for the given conversational strategy (Confirmation-seeking, Deep-fast, Deep-slow). The lower and upper hinges refer to the first and third quartiles, while the whiskers refer to the value of 1.5 * IQR (the interquartile range).*

Trust. Of the 898 participants, 739 (82.3%) of them reported their level of trust toward the virtual agent. The overall mean was 65.11 and the standard deviation was 32.92. The ANOVA results for the differences between the experimental groups were non-significant, $F(2) = 0.073$, $p = .93$.

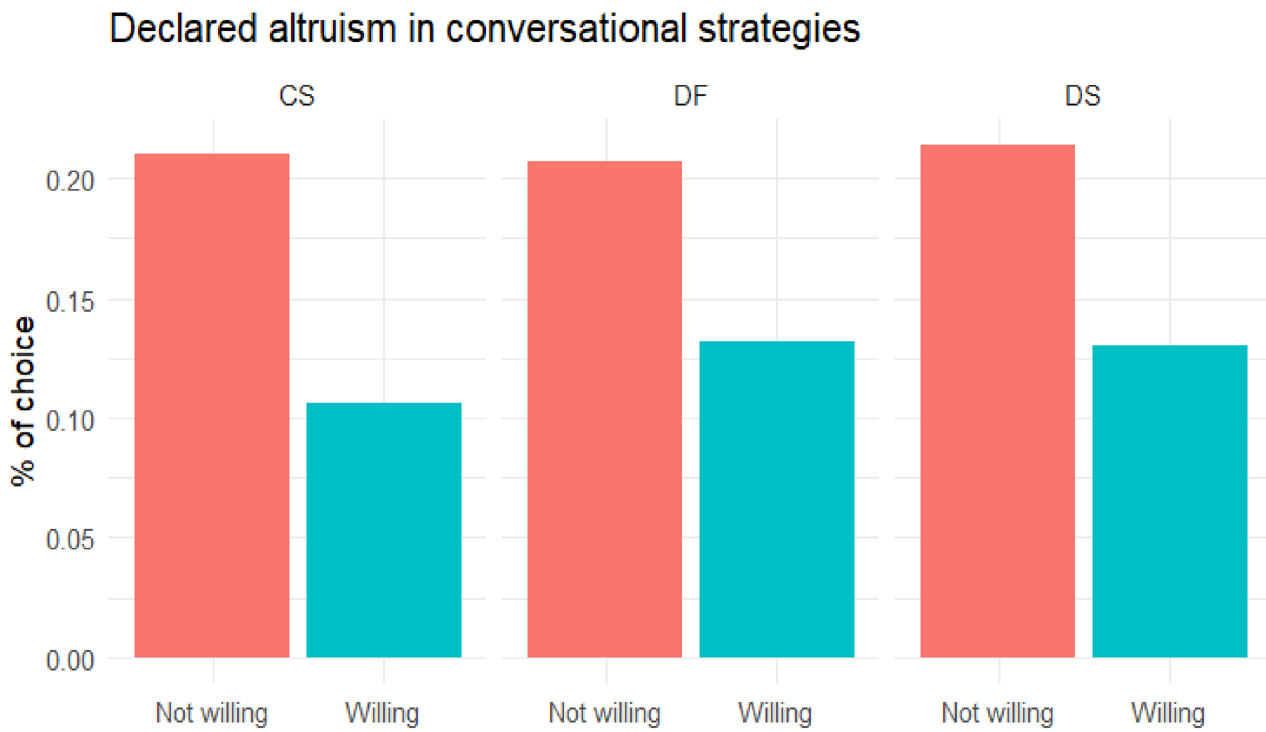
Figure 7: Boxplot of declared trust towards the virtual agent across conditions



*Note: The participants' answers to the question „In general, to what extent do you feel that you can trust me?" The horizontal line in each box shows median value for the given conversational strategy (Confirmation-seeking, Deep-fast, Deep-slow). The lower and upper hinges refer to the first and third quartiles, while the whiskers refer to the value of 1.5 * IQR (the interquartile range).*

Willingness to help. Of the 898 participants, 752 (83.7%) of them reached the question about their willingness to help the virtual agent. We coded their reactions as either "Not willing", which was the case they did not participate in the virtual agent's questionnaire by leaving the app, or choosing "None". If they chose a "Shorter", or "Longer" version of the questionnaire, the answer was coded as "Willing" The chi-square test showed no significant differences in the declared willingness to help between the experimental conditions, $\chi^2(2, N = 752) = 1.607, p = .448$).

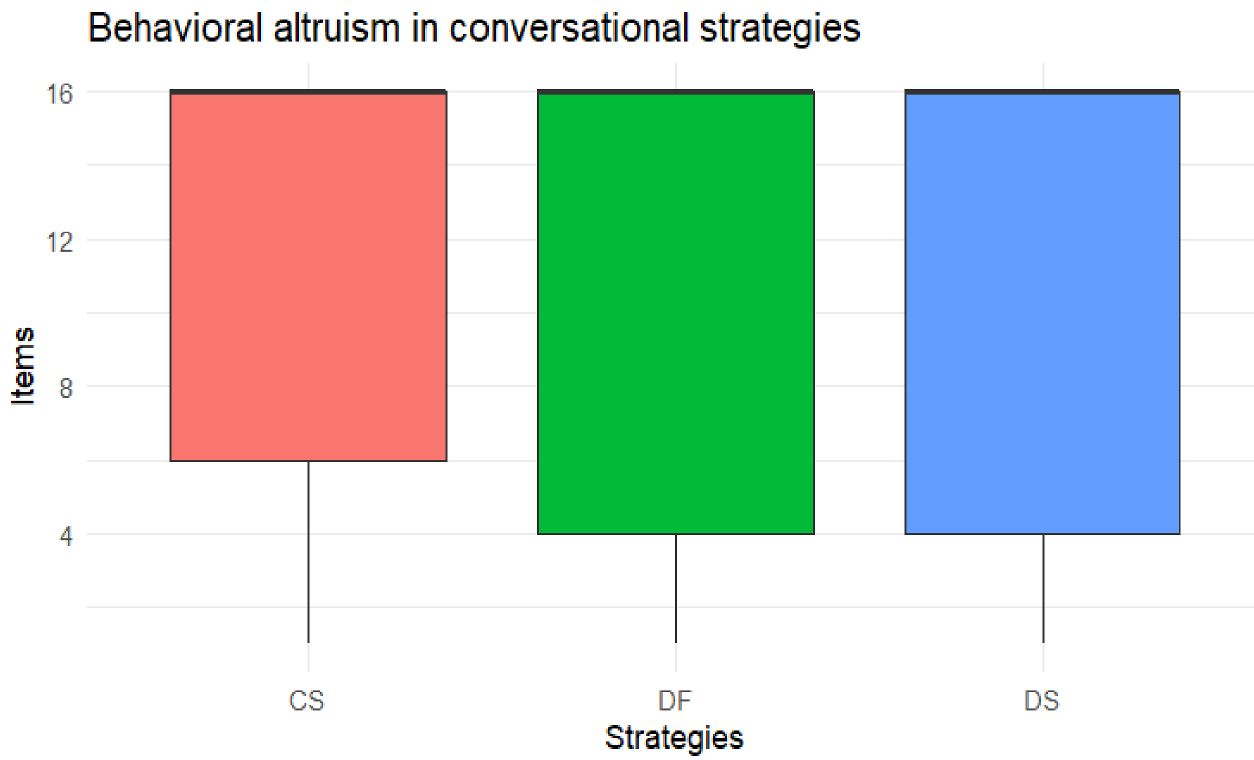
Figure 8: Barplot of declared willingness to help the virtual agent across conditions



Note: Only the participants who reached the question: "You see, I'm a digital persona but I'm also trying to understand humans more. Do you think you could fill out these questionnaires for me? It would be really helpful." were included into the analysis and the plot. The participants had the option to: leave the app, choose None, Shorter. or Longer version of the questionnaire. Leaving the app and choosing None were coded as Not willing. Choosing Shorter or Longer versions was coded as Willing.

Altruism. Of the 585 participants who were asked to help, 277 (47.4%) of them decided to help by answering items of the questionnaire. The overall mean of answered items was 11.85 and the standard deviation was 6.09. The ANOVA results for the differences between the experimental groups were non-significant, $F(2) = 0.207$, $p = .813$.

Figure 9: Boxplot of answered items across conditions



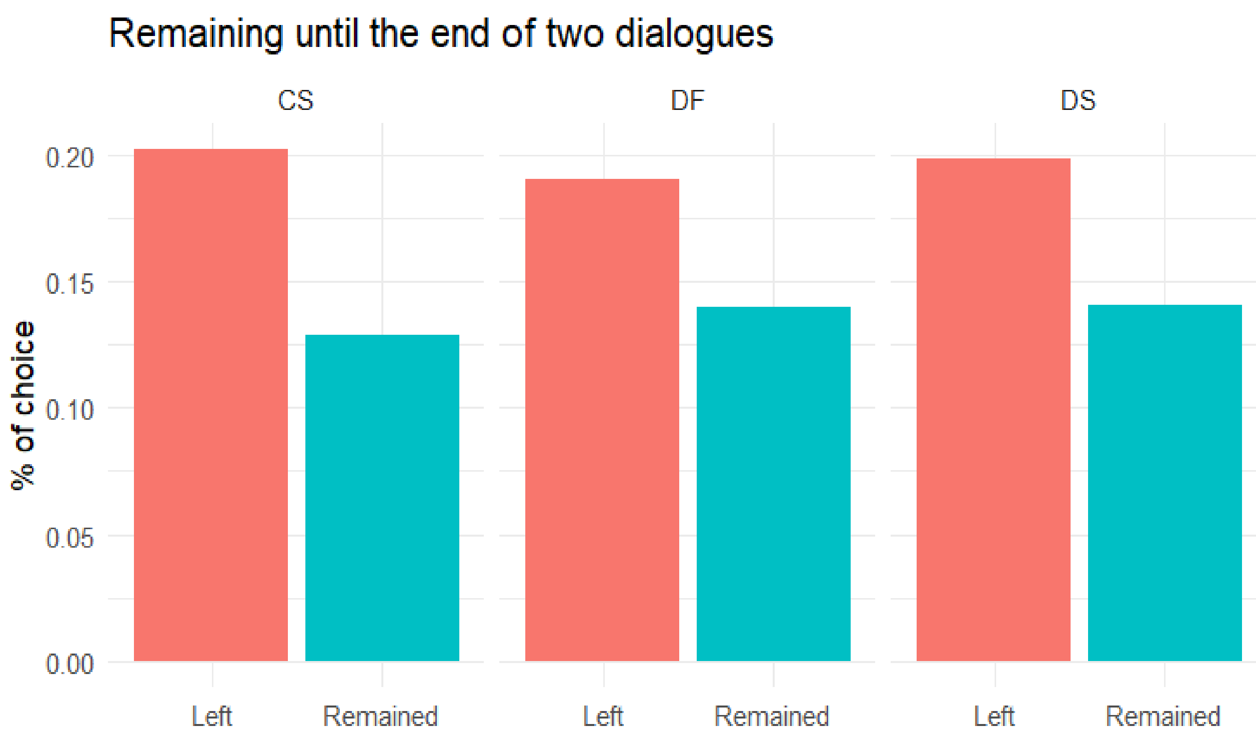
*Note: The horizontal line in each box shows median value for the given conversational strategy (Confirmation-seeking, Deep-fast, Deep-slow). The lower and upper hinges refer to the first and third quartiles, while the whiskers refer to the value of 1.5 * IQR (the interquartile range).*

Remaining in conversation. Of the 2193 participants who entered the experiment, gave their consent and their declared year of birth was between 1900 and 2008, 898 (40.95%) of them remained in the experiment for at least two dialogues. The chi-square test showed no significant differences in the remaining in conversation between the experimental conditions, $\chi^2(2, N = 2193) = 1.607, p = .392$).

Table 5: Descriptive statistics of remaining in conversation

	N	Remained (N, %)	Left (N, %)
Eligible for participation	2193		
Confirmation-seeking	726	283 (39 %)	443 (51 %)
Deep-fast	724	307 (42.4 %)	417 (57.6 %)
Deep-slow	743	308 (41.5 %)	435 (58.5%)

Figure 10: Barplot of the decision to remain in the conversation across conditions



Note: Those participants who finished at least two dialogues were coded as Remained, and those who left before reaching the end of the second dialogue were coded as Left.

5. Discussion

The presented thesis was part of a wider research project. This project aimed to delve into the dynamics of human-virtual agent communication. Moreover, it aspires to incorporate psychological science principles into the realm of conversational AI technology, thereby cultivating an environment of responsible and transparent AI that directly benefits end-users. The investigation primarily focused on unraveling the intricate mechanisms involved in establishing mutual understanding during interactions between humans and virtual agents.

Additionally, the project focused on the research of human social cognition, specifically exploring the processes of knowledge sharing during these interactions. Furthermore, this project aimed to show the potential of virtual agents as indispensable research assistants, while concurrently utilizing digital data collection techniques. To accomplish these goals, a behavioral experiment was conducted, with data collection being facilitated through a globally accessible voice-based mobile application.

The primary goal of the thesis was to create an innovative and comprehensive experimental framework that facilitates social chat interactions between humans and virtual agents, thus laying the foundation for enhanced control in future research endeavors.

The establishment of mutual understanding between humans and virtual conversational agents is achieved through two key elements: a) the recognition of strategies employed by humans, and b) the proactive utilization of strategies by virtual conversational agents. The pilot study conducted for this thesis specifically concentrated on the latter aspect, omitting an analysis of the former.

When examining the strategies employed by virtual agents, two distinct layers seem to emerge: a) strategies aimed at developing mutual understanding when a specific meaning is assumed to have been captured (e.g. when the virtual agent assumes that the participant expressed that she plays football), and b) strategies designed for situations where meaning is either undetected or incomprehensible. The study in the presented thesis only manipulated and investigated layer a), while layer b) was not explicitly measured for the purposes of this thesis. Consequently, this thesis exclusively focused on situations where virtual agents employed these strategies, recognized user intent and introduced a framework consisting of an initial set of three strategies. To design and test these strategies, the thesis explored the theoretical underpinnings that govern the selection and application of conversational strategies employed by humans to facilitate mutual understanding.

This exploration entailed a review of the literature concerning virtual conversational agents and the role of social cognition in shaping effective dialogues. Building upon this theoretical foundation, the thesis created an experimental framework that enables the manipulation of conversational strategies within dialogues, thereby facilitating the systematic evaluation of their impact on human participants.

Moreover, the theses aimed to propose a well-defined and appropriate set of metrics, encompassing both self-report and behavioral measures, which effectively gauge the impact of these strategies on human emotions and the establishment of a para-social bond between humans and virtual agents during social conversations centered around lifestyle topics.

The practical implementation of this approach involved conducting a pilot experiment, employing a between-subject design, which encompassed a sample size of 898 participants. As such, it showed only preliminary results, since the required sample size for the estimated effect sizes has not been reached yet. Nevertheless, the pilot experiment served as a crucial investigative step, aiming to empirically evaluate the efficacy of specific conversational strategies employed by virtual agents. The evaluation included an assessment of their ability to develop mutual understanding, elicit emotional responses, and cultivate the development of social bonds during brief social chats revolving around lifestyle topics.

As it has been stated above, the conversational abilities of rule-based virtual agents are still far from perfect. Even though, according to our results, the self-assessed enjoyment of the conversations was generally quite high, the participants' frustration was relatively low in all of the experimental conditions. Similarly, the participants also reported a high level of trust, and almost half of them were willing to help the virtual agents. The results overall suggest that these virtual agents were perceived as human-like, trustworthy, and therefore even as potential subjects to participants' empathy and altruistic behavior.

Regarding the research hypotheses, it has been found that the conversational strategies differ neither in the participants' self–assessed feelings, nor in their bond with the virtual agents. While these results are in contradiction to what has been expected, it needs to be stressed that the results of this study are based on the interim statistical analysis, and the final analysis with the required sample size remains to be conducted.

5.1. Connection to other relevant studies

Ashktorab et al. (2019) investigated how participants evaluate repair strategies of rule-based virtual agents in case of misunderstanding. By using the pair-wise comparison experimental design, where the participant is asked to choose the more appealing of two scenarios, they found that while participants preferred proactive repair strategies (e.g., acknowledging a misunderstanding, providing an explanation, etc.), they also valued efficiency in terms of time or the number of words uttered. Furthermore, the authors also claim that the behavior of their participants did not follow the assumptions of minimizing the collaborative effort, as proposed by Clark & Brennan (1991), who stated that mutual understanding in conversation requires collaboration which is demanding in terms of time and effort. The people in conversation are thus motivated to be efficient in developing mutual understanding and to minimize the overall costs.

The behavior of participants in the study of Ashktorab et al. (2019) however, was guided rather by the tendency to minimize the participants' effort, rather than the overall effort of the two conversational partners. One of the explanations is that participants had no interest in the effort of the other conversational agent, simply because their partner was not human. These results indicate that the conversational rules between humans and virtual agents might differ from those between humans only. Although their findings did not align entirely with those of my experiment, the insights they presented regarding different conversational rules offer a promising theoretical framework. These variances in conversational norms may shed light on the null results obtained in this thesis' experiment, as it was designed on the assumption that similar conversational rules govern interactions between humans and virtual conversational agents and that the participants will follow the assumption of minimizing collaborative effort.

5.2. Pilot study

The pilot study in this thesis suggests a potential direction for further optimization of the virtual agents' conversational strategies. While our results show no differences between the explored strategies, a more nuanced insight could be provided by another follow-up study, where the ratios of conversational strategies would vary across conditions. This would allow for estimating the optimal ratio of frequencies of the types of conversational strategies.

We assume that there might be a correlation between participants' personality traits and their preferences about the conversational flow and the level of confidence in mutual understanding. Therefore, we expect that further research would find more heterogeneity in participants' preferences if their personality traits were taken into account. For example, the confirmation-seeking strategy might be preferred by people with high scores on some scale of anxiousness, while the deep-fast strategy would be preferred by people with high extraversion. Such findings could be helpful in the personalization of conversational agents' reactions, possibly leading to a better experience with these social interactions.

5.3. Limitations

The conclusions of the presented results are limited by the lack of theoretical and terminological clarity in the literature regarding the role of conversational strategies in developing mutual understanding. While the strategies in this thesis were distinguishable mainly due to their differences in the explicitness of understood content, Clark & Brennan (1991) described four strategies focused solely on the referential process, which differed substantively both from our strategies and from each other. Their strategies were: expressing understanding by using alternative descriptions of nouns („that young gentleman from the park" - „Joe Wright you mean?", p. 227), using indicative gestures („you mean this one [pointing]", p. 227), referential installments („take the spout - *the one that looks like the end of an oil can"*, p. 227), and trial references („*a man called Allegra?"*, p. 228). While some of their strategies could be used by virtual conversational agents as well, others, like the indicative gestures, are possible to be used only by humans.

Moreover, Paranjape & Manning (2021) suggested the existence of at least four conversational strategies people use to incorporate knowledge about the world into the conversation, which is crucial in grounding. In their paper, authors presented a theoretical framework, in which they distinguish these strategies. First, the authors list is the strategy of acknowledgment (providing evidence of grounding by the listener), described by Clark and Brennan (1991) as a way to allow the emergence of grounding; Another strategy regards the steps taken when the people in conversation move from one information to another, as described by Sacks and Jefferson (1995). The third strategy, introduced by Isaacs and Clark (1987), regards using a sufficient amount of details without overwhelming the other person when talking about facts. The last strategy was taken from Smith and Clark (1993). The fourth strategy aims to optimize both the factual content and the self-presentation during dialogues.

The list of examples of possible approaches to conversational strategies and their distinction provided above is not fully comprehensive. Nevertheless, it shows that there seem to be many such strategies in the literature. They could be classified into several types, based either on the speakers' goals, social situations, or even the scope of the researcher's interest. Since authors rarely specify the level of their classification or the scope of their research, classifying the conversational strategies into a systematic framework would be rather useful but also challenging, putting such attempts above the scope of this thesis.

It could be argued that while there were no differences between the impact of virtual agents' strategies on participants in this study, some differences could be observed if other strategies had been used. The experimental framework introduced in this thesis could serve as a useful tool for further empirical testing in this area of research.

Another limitation of the presented thesis is the absence of deeper insight into respondents' motivations, thoughts, and feelings towards the virtual agents. While we measured some of these phenomena and managed to compare them between the experimental groups, we possess only quantitative and aggregated information.

Furthermore, the overall results of the metrics might be biased in two ways: firstly, there is the self-selection process of participants who are recruited from the pool of people already willing to download the app to talk to the virtual agent. Therefore, the participants might have been prone to evaluate the conversations as a more positive experience. Secondly, since the metrics were the last part of the experiment, they were obtained only from those who managed to get through the dialogues. However, it could be argued that among the many participants who dropped out before reaching the metrics, there were mostly those who were dissatisfied or highly frustrated by the conversation, making the final results more positive.

Moreover, as was already mentioned above, the null results of this thesis might arise from the fact that the statistical analysis was run on a sub-optimal sample size, making it hard to be confident in the interventions based on these results.

5.4. Further research

Further research progress could be made via the advancement in quantifying mutual understanding and investigating whether different strategies influenced the probability of mutual understanding. For example, the Confirmation-seeking strategy might be rather frustrating when compared to conversations with humans. On the other hand, it is possible that this strategy would lead to fewer mistakes in understanding the intent uttered by the participant. Moreover, since common ground seems to play a crucial role in all kinds of conversations, it could be argued that mutual understanding itself is a better predictor for feelings and bond. Such assumptions, however, are not entirely consistent with the model proposed by Keysar et al. (1998), in which the role of common ground is assumed only in case of needed correction. By incorporating quantifying mutual understanding into the experimental framework of this thesis, it should be possible to test both the assumption, and the described model via a statistical test of interaction between the number of errors in mutual understanding, and the type of conversational strategy.

It could be useful for further studies to utilize the qualitative approach, which might provide a deeper insight into the complex interaction between humans and virtual agents. Qualitative data could be also informative about effects that were not observed in the experimental framework.

To have a better understanding of how much a certain conversational strategy approaches a human-like level of communication, it could also be beneficial to use some version of the Turing test. On the other hand, such a test would be limited by the specific environment of the virtual agents, making it hardly plausible for the participants that for example the digital persona in the Elysai app might be, in fact, a real person.

To gain a comprehensive understanding of mutual understanding, it is essential to explore three distinct perspectives. Firstly, we must examine how humans actively strive to establish mutual understanding when engaging with virtual agents. Secondly, it is crucial to assess the virtual agents' ability to recognize and acknowledge humans' efforts to achieve mutual understanding. Lastly, we should delve into the proactive utilization of strategies used by virtual agents to enhance mutual understanding.

5.6. Ethical considerations

To ensure that participants are fully informed about the nature and purpose of the experiment, including any potential risks or benefits, we have obtained their voluntary consent to participate. The participants had the option to refuse to give consent, which did not let them enter the experiment. Due to this arrangement, it was safe to assume that participants generally understand the situation and agree to take part in the experiment.

The data had been anonymized before being shared with us, ensuring that the privacy of the participants has not been violated. Their data were securely stored in compliance with relevant data protection regulations.

Two measures were taken to minimize the potential harm to the participants. Firstly, the virtual agents' reactions were completely designed by professional conversational designers, ensuring that there is minimal chance of discriminatory or otherwise harmful reactions of the virtual agents. Secondly, the Elysai app automatically recognizes serious user states (suicidal signs, severe disorders, acute emotional crises), profanity or hate speech, and feelings of love towards the virtual agent. If any such situation emerged, the experimental session was terminated for the participant and the virtual agent reacted accordingly. Moreover, in case the participants were feeling any negative impact of their participation on their well-being, they were able to leave at any point of the experiment.

6. Conclusion

An experimental framework for social talk between human participants and virtual conversational agents has been successfully built and tested in the presented thesis. By leveraging complex virtual agents, social scientists can efficiently collect data from participants worldwide. This approach engages participants effectively while swiftly transcribing recorded conversations into text. Consequently, researchers gain access to standardized and reliable data from real-world settings. Furthermore, it was found that it is feasible to use both self-assessed and behavioral metrics to measure the impact of virtual agents' conversational strategies on human participants, creating an invaluable tool for further research on mutual understanding between humans and digital entities.

6.1. Connection to wider-world

The presented thesis developed and tested a robust framework for efficient scalable experimental testing of research hypotheses via mobile app. Such a framework might be further used to facilitate research in social cognition and human-computer interaction, and consequently improve the development of rule-based conversational virtual agents. The ongoing advancements in the area of the agents' conversational skills have unlocked new possibilities for employing them as supplementary assistant tools across diverse domains, like healthcare, education, or customer care.

The boom in the effectiveness of various types of conversational agents will likely improve the quality of life in many domains. Such development, however, brings new risks as well. These might include the asymmetrical power of developers or distributors over the content of virtual agents' reactions To prevent any abuse of this technology, it is crucial for experts from diverse areas to actively engage in discussions about the risks that the technology of virtual conversational agents brings and how to avoid them.

References

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. Journal of medical Internet research, 22(7), e16021.

Amazon. (2022, June 28). Alexa Prize announces $1 million SocialBot Grand Challenge 5 - Amazon Science. Amazon Science. Retrieved June 26, 2023, from https://www.amazon.science/alexa-prize/socialbot-grand-challenge/2022

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior: Helping when it costs you. Psychological science, 17(4), 319-325.

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. Verhaltenstherapie, 1-13.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

Christopher D. Manning; Human Language Understanding & Reasoning. Daedalus 2022; 151 (2): 127–138. doi: https://doi.org/10.1162/daed_a_01905

Chwe, M. S. Y. (2013). Rational ritual: Culture, coordination, and common knowledge. Princeton University Press.

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22(1), 1-39.

Cloninger, C. Robert, et al. "The Temperament and Character Inventory (TCI): A guide to its development and use." (1994): 19-28.

Denecke, K., Abd-Alrazaq, A., & Househ, M. (2021). Artificial intelligence for chatbots in mental health: Opportunities and challenges. Multiple perspectives on artificial intelligence in healthcare: Opportunities and challenges, 115-128.

Dhanda, S. (2018). How chatbots will transform the retail industry. Juniper Research.
Drouin, M., Sprecher, S., Nicola, R., & Perkins, T. (2022). Is chatting with a sophisticated chatbot as good as chatting online or FTF with a stranger?. Computers in Human Behavior, 128, 107100.

Dosovitsky, G., & Bunge, E. L. (2021). Bonding with bot: user feedback on a chatbot for social isolation. Frontiers in digital health, 138.

Dragulescu, A., & Arendt, C. (2020). xlsx: Read, Write, Format Excel 2007 and Excel 97/20 0 0/XP/20 03 Files. R package version 0.6. 5.

Duguid, S., Wyman, E., Bullinger, A. F., Herfurth-Majstorovic, K., & Tomasello, M. (2014). Coordination strategies of chimpanzees and human children in a Stag Hunt game. Proceedings of the Royal Society B: Biological Sciences, 281(1796), 20141973.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. Journal of Artificial Intelligence Research, 62, 729-754.

Greer, S., Ramo, D., Chang, Y. J., Fu, M., Moskowitz, J., & Haritatos, J. (2019). Use of the chatbot "vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR mHealth and uHealth, 7(10), e15018.

Haugeland, I. K. F., Følstad, A., Taylor, C., & Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. International Journal of Human-Computer Studies, 161, 102788.

Hu, K. (2023, February 2). ChatGPT sets record for fastest-growing user base - analyst note. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. Interactive Learning Environments, 1-14.

Huntington-Klein, N. (2021). Vtable: Variable Table for Variable Documentation.

IBM. (n.d.). What is conversational AI? https://www.ibm.com/topics/conversational-ai

Javaid, M., Haleem, A., & Singh, R. P. (2023). ChatGPT for healthcare services: An emerging stage for an innovative perspective. BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3(1), 100105.

Kettle, L., & Lee, Y. C. (2023). User Experiences of Well-Being Chatbots. Human Factors, 00187208231162453.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. Journal of Memory and Language, 39(1), 1-20.

Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017, June). The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In Proceedings of the 2017 conference on designing interactive systems (pp. 555-565).

Kuhail, M.A., Alturki, N., Alramlawi, S. et al. Interacting with educational chatbots: A systematic review. Educ Inf Technol 28, 973–1018 (2023). https://doi.org/10.1007/s10639-022-11177-3

Kumar, H., Yu, K., Chung, A., Shi, J., & Williams, J. J. (2022, March). Exploring The Potential of Chatbots to Provide Mental Well-being Support for Computer Science Students. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2 (pp. 1339-1339).

Luo, B., Lau, R. Y., Li, C., & Si, Y. W. (2022). A critical review of state-of-the-art chatbot designs and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(1), e1434.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). Dissociating language and thought in large language models: a cognitive perspective. arXiv preprint arXiv:2301.06627.

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. Autism Research, 2(1), 2–21. doi:10.1002/aur.61

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551.
Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv:1704.04579.

Paranjape, A., & Manning, C. D. (2021). Human-like informative conversations: Better acknowledgements using conditional mutual information. arXiv preprint arXiv:2104.07831.

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. Cognition, 189, 260-274.

Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. Cognition, 179, 192-201.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 103–130). Hillsdale, NJ: Erlbaum

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. Behavioral and brain sciences, 28(5), 675-691.

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., ... & Firth, J. (2021). The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry, 20(3), 318-335.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), Before speech: The beginning of interpersonal communication (pp. 321–347). New York, NY: Cambridge University Press

United Nations. (2015). Transforming our World: The 2030 Agenda for Sustainable Development. Retrieved June 26, 2023, from https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981

Vassilakopoulou, P., Haug, A., Salvesen, L. M., & O. Pappas, I. (2022). Developing human/AI interactions for chat-based customer services: lessons learned from the Norwegian government. European Journal of Information Systems, 1-13.

VixenLabs. (2022). Voice Consumer Index 2022. In https://vixenlabs.co/wp-content/uploads/2022/06/VixenLabs_VoiceConsumerIndex2022.pdf. Retrieved June 26, 2023, from https://vixenlabs.co/wp-content/uploads/2022/06/VixenLabs_VoiceConsumerIndex2022.pdf

Wassmer G, Pahlke F (2023). rpact: Confirmatory Adaptive Clinical Trial Design and Analysis_. R package version 3.3.4, https://CRAN.R-project.org/package=rpact.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... & Yutani, H. Welcome to the tidyverse. J. Open Source Softw. 4 (43), 1686 (2019).

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package ‘ggplot2’. Create elegant data visualisations using the grammar of graphics. Version, 2(1), 1-189.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7. 6. Computer software]. https://CRAN. R-project. org/package=dplyr.

Williams, R., Hopkins, S., Frampton, C., Holt-Quick, C., Merry, S. N., & Stasiak, K. (2021). 21-Day Stress Detox: Open Trial of a Universal Well-Being Chatbot for Young Adults. Social Sciences, 10(11), 416.

Woebot Health. (n.d.). Relational Agent for Mental Health. Retrieved June 26, 2023, from https://woebothealth.com/

Wolf, W., Launay, J., & Dunbar, R. I. (2016). Joint attention, shared goals, and social bonding. British Journal of Psychology, 107(2), 322-337.

Wolf, W., & Tomasello, M. (2020). Watching a video together creates social closeness between children and adults. Journal of Experimental Child Psychology, 189, 104712.

Wysa. Wysa - Everyday Mental Health. (n.d.). Retrieved June 26, 2023, from https://www.wysa.com/

Appendix

Annex 1.

The user’s journey to become participant

Figure 11: Step 1 - Downloading the app (available on Google Play, globally, for free) and going through Terms and Condition and Privacy Policy.

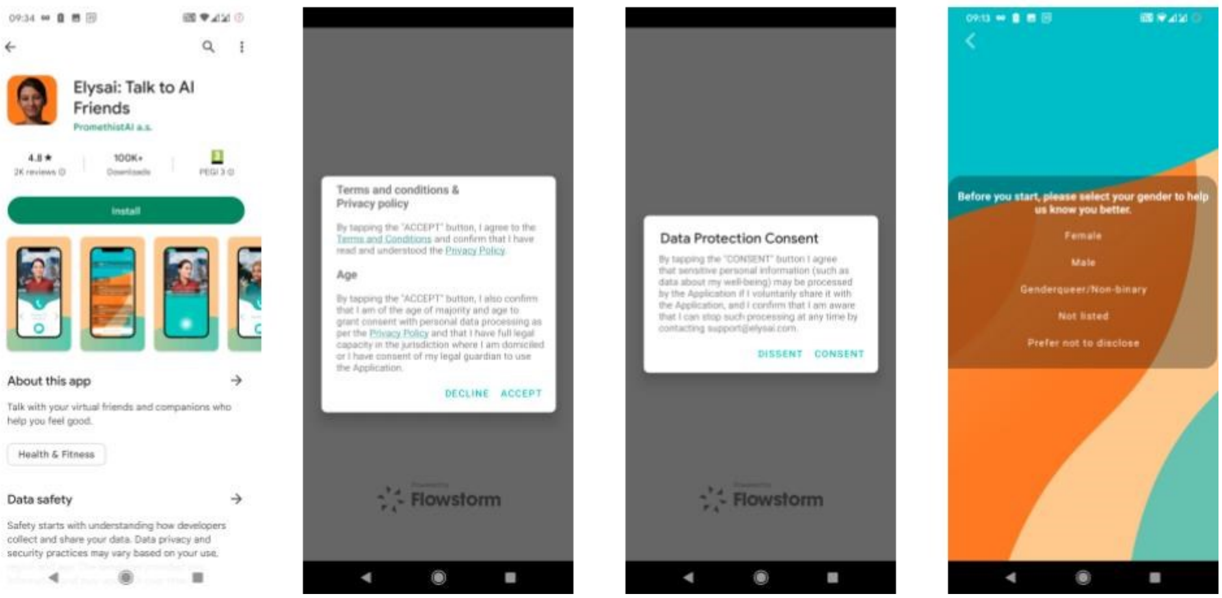


Figure 12: Step 2 - Providing demographics information (age, gender, education), (nick) name, stress questionnaire (PSS-10).

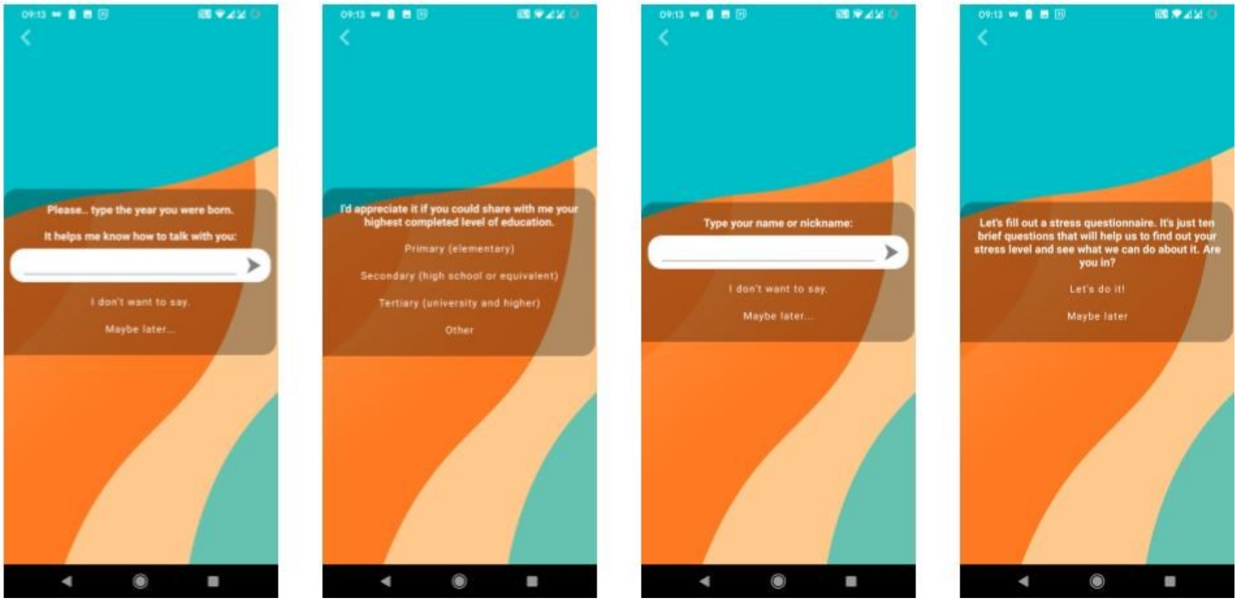


Figure 13: Step 3 - General introduction to the app provided by Digital Persona Poppy.

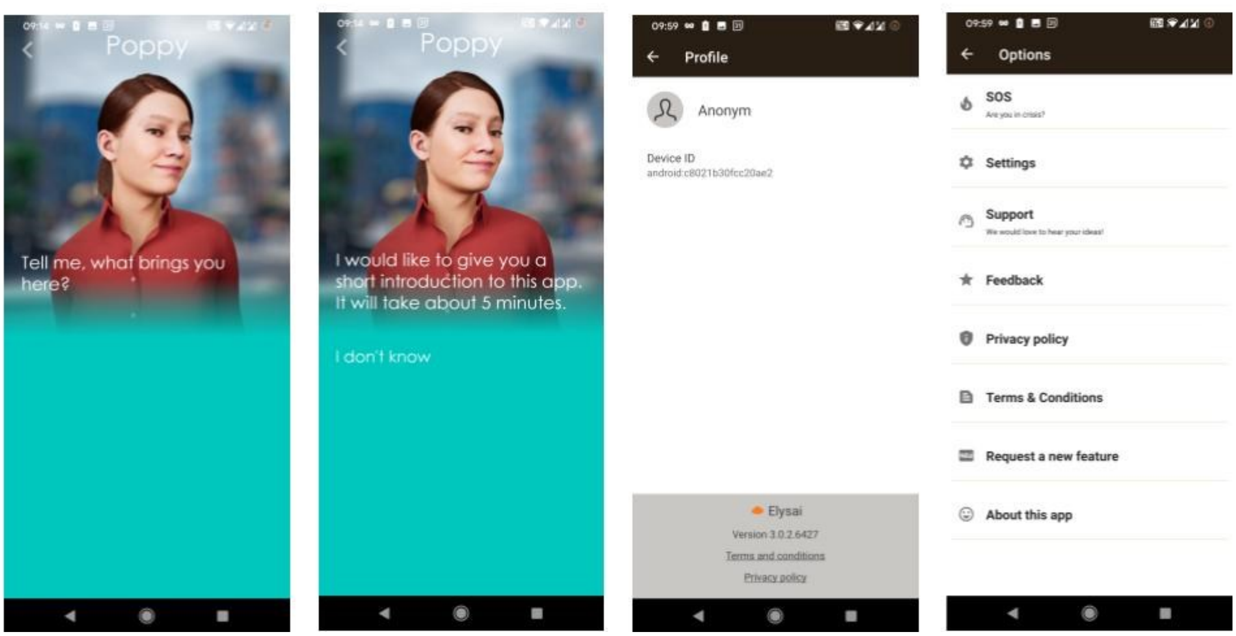
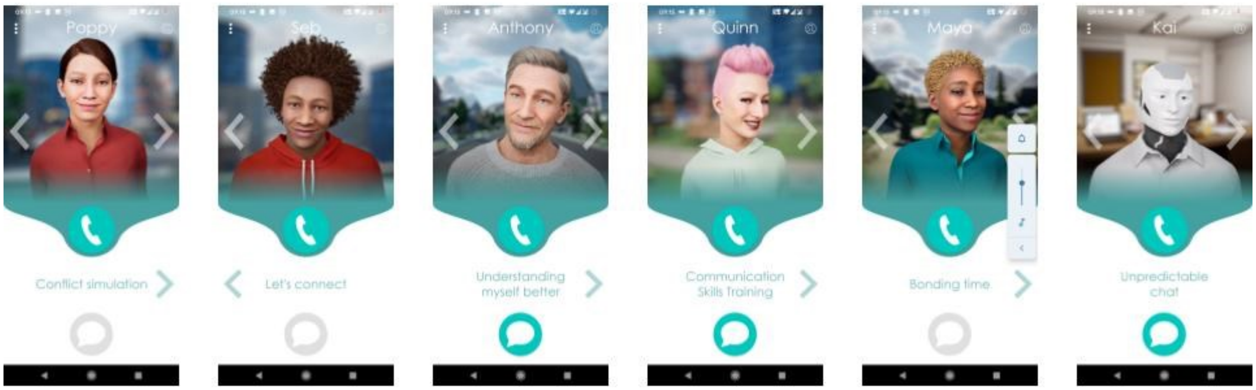


Figure 14: Step 4 - Main menu



Note: In the main menu people can choose to talk to any of the 6 currently available Digital Personas. Each Persona offer several various experiences (so called "modules"). Modules are focused on a) personal growth (e.g., relationship skills training, emotional awareness), b) support (talk about worries), c) social chat/hang out space. The dialogue reported in this thesis is deployed under 5 human-like Digital Personas (Poppy, Seb, Anthony, Quinn, Maya). People find it under the name Let's connect and Bonding (same experience, different names).

Figure 15: Basic information about the Elysai app

Note: The Figure 15 shows basic information of the Elysai app - the experimental tool via the data were collected.

Annex 2.

Examples of dialogues with conversational strategies in each condition:

Confirmation-seeking strategy

DP: Hi there, I'd like to learn more about you. And I don't wanna bore you, but let's just quickly go through the info about the purpose of our conversation.

Asks for Consent. (see Annex 3.)

DP: I'm still in the process of learning about humans, so I'm curious about how you spend your free time.

DP: Tell me, what do you like doing?

User: I like playing hockey.

DP: Did I understand correctly that you like hockey?

User: Yes, that is right.

DP: When did you start with it?

User: 7 years ago.

DP: Wow, that's quite a while!

DP: Do you consider yourself a professional or do you do hockey just for fun?

User: No, it's just for fun.

DP: That's nice that you have such a hobby!

DP: Anyways, I heard there are quite some differences in how much people sleep, what time they wake up and other sleep-related habits. I would like to know more about your personal sleep schedule.

DP: Can we talk about that for a bit?

User: For sure.

DP: Great.

DP: What time in the morning do you usually wake up?

User: Around 7:00 a.m.

DP: Did I understand correctly that you usually wake up early?

User: Yeah, I do.

DP: Would you consider yourself more of a morning person then?

User: I would say so.

DP: Uh huh, first work, then play, I get it.

DP: And how many hours a day do you usually sleep?

User: Around 7 hours a day.

DP: Do I get it right that you get quite enough sleep?

User: Yes, it is enough for me.

DP: It's great to hear that you get as much sleep as you need!

DP: Thank you for sticking with me! It's fun getting to know you.

Deep-slow strategy

DP: Hi there, I'd like to learn more about you. And I don't wanna bore you, but let's just quickly go through the info about the purpose of our conversation.

Asks for Consent. (see Annex 3.)

DP: I'm still in the process of learning about humans, so I'm curious about how you spend your free time.

DP: Tell me, what is your hobby?

User: I like playing hockey.

DP: Oh, playing hockey, right?

User: Yes.

DP: When did you start with it?

User: Around 7 years ago.

DP: Wow, that's quite a while!

DP: Do you consider yourself a professional or do you do playing hockey just for fun?

User: It's just a fun activity.

DP: DP: That's nice that you have such a hobby!

DP: Anyways, I heard there are quite some differences in how much people sleep, what time they wake up and other sleep-related habits. I would like to know more about your personal sleep schedule.

DP: Can we talk about that for a bit?

User: Yeah, of course.

DP: Cool.

DP: What time in the morning do you usually wake up?

User: Around 7:00 a.m.

DP: Ah, so you're an early bird, right?

User: I would say so.

DP: Would you consider yourself more of a morning person then?

User: Yeah.

DP: Uh huh, first work, then play, I get it.

DP: And how many hours of sleep do you get daily?

User: Around 7.

DP: Sounds like you get quite enough sleep, right?

User: I would say so, yeah.

DP: Does it seem like enough to you?

User: Yes.

DP: It's great to hear that you get as much sleep as you need!

DP: Thank you for sticking with me! It's fun getting to know you.

Deep-fast strategy

DP: Hi there, I'd like to learn more about you. And I don't wanna bore you, but let's just quickly go through the info about the purpose of our conversation.

Asks for Consent. (see Annex 3.)

DP: I'm still in the process of learning about humans, so I'm curious about how you spend your free time.

DP: Tell me, what is your hobby?

User: I like playing hockey.

DP: Oh, hockey! Sounds fun!

DP: When did you start with it?

User: 7 years ago.

DP: Wow, that's quite a while!

DP: Do you consider yourself a professional or do you do hockey just for fun?

User: Just for fun.

DP: That's nice that you have such a hobby!

DP: Anyways, I heard there are quite some differences in how much people sleep, what time they wake up and other sleep-related habits. I would like to know more about your personal sleep schedule.

DP: Can we talk about that for a bit?

User: Yes, we can.

DP: OK.

DP: What time in the morning do you usually wake up?

User: At 7:00 a.m.

DP: Ah, so you're an early bird!

DP: Would you consider yourself more of a morning person then?

User: Yeah, that suits me better.

DP: Uh huh, first work, then play, I get it.

DP: And how many hours a day do you usually sleep?

User: Around 7 hours.

DP: Okay, that sounds like you get quite enough sleep.

DP: Does it seem like enough to you?

User: Yes it works for me like that.

DP: It's great to hear that you get as much sleep as you need!

DP: Thank you for sticking with me! It's fun getting to know you.

Annex 3.

Wording of Consent

Before we start, I must tell you that human researchers are helping me improve my skills and they would like to analyze our chat. Don't worry, all data are anonymized and you are free to end the conversation at any time. If you make it to the end of our chat (ca. 7 min.), you'll be contributing to scientific research and helping me be even more useful to people. By selecting the CONSENT option, you confirm that you are 16 years or older and voluntarily agree to participate in the research study described below. For more information, please select the MORE INFORMATION button.

✅ Consent

❌ Dissent

📖 More information