Breaking the Silence:
Encouraging Conversations with Digital Humans

MASTER THESIS

Breaking the Silence: Encouraging Conversations with Digital Humans

Master Thesis

Authors:

Rebecca Damgaard Hansen

Sofie Kristine Thøgersen

Jeppe Firring Givskud

Master Thesis - Engineering Psychology

Study group: 1086 - PDP10 - Digital Humans

Supervisor: Rodrigo Ordoñez

Semester theme: Master Thesis

AALBORG UNIVERSITET

STUDENTERRAPPORT

February 1st - june 4th 2025

Number of pages: 135 **Attachments: 10** **Appendix: 6** **ECTS:** 30 ECTS

Abstract

As digital humans increasingly find application in service-oriented environments, understanding how users engage with this new technology becomes critical to improve their usability and adoption. This master's thesis investigates how conversational initiation, whether initiated by the user or the AI agent influences user engagement and experience. Drawing on literature and field studies at Aarhus Jobcenter and Krifa, the study identifies barriers to interaction with digital receptionists and highlights the need for clearer social cues and interaction affordances.

To explore these dynamics, a digital human named Rosie was deployed in semi-public spaces at Aalborg University across two experimental conditions: one in which Rosie proactively initiated conversations, and one where she remained passive. In total there were 175 trials in which 29 participants interacted with the AI. The interaction was investigated through probability of success, real-time observations, exit interviews, and questionnaires asking about two constructs: the Technology Acceptance Model (TAM) and Temple Presence Inventory (TPI). The study found no statistically significant difference in either probability of success, TAM or TPI.

However, users in the Rosie-initiated condition were significantly less likely to request facilitator help, suggesting that proactive initiation reduces ambiguity by providing a strong signifier to voice interaction. Thematic analysis further revealed that user initiation shaped users' subjective impressions, eliciting more reflective and cognitively engaged responses.

These findings suggest that conversational initiation plays a pivotal role in shaping interpretive aspects of human–AI interaction. Designing digital humans to proactively initiate dialogue can reduce confusion on how to interact with the system which may enhance user engagement, reduce social friction, and potentially improve perceived intelligence, particularly in socially situated contexts. The study underscores the importance of interaction framing and highlights the value of continued research into the contextual and social dynamics of AI-mediated encounters.

Chapter 1 - Preface

This master's thesis was developed in collaboration with NTT Data Business Solutions regarding their product *digital humans*.
The project group would like to thank Rasmus Gade Ørtoft from NTT Data for help, discussions and feedback during the project.
Our deepest gratitude goes to our supervisor, Rodrigo Ordoñez, for his dedicated guidance and continuous encouragement during the entire process.

This master's thesis explores the concept of digital humans. While the term can encompass a wide range of technologies, in this context it specifically refers to voice interfaces that feature a human-like face and are capable of engaging in spoken interaction. These systems are also commonly referred to as embodied conversational agents (ECAs), virtual humans, avatar-based interfaces, or multimodal dialogue agents.

Such agents combine speech synthesis, facial animation, and sometimes gesture or gaze behavior to create the impression of a human-like presence. They are increasingly used in fields such as customer service, healthcare, education, and entertainment, offering more natural and engaging ways for users to interact with technology.

The digital human used for testing in this study is named Rosie. For the sake of readability and narrative clarity, Rosie is occasionally referred to using the pronouns she/her. This choice is made purely for stylistic and linguistic convenience and does not imply any assumptions about the autonomy, identity, or personhood of digital humans or artificial intelligence.

The report uses APA style for references.
Both appendixes and attachments are referred to by name.

Contents

Chapter 2 - Introduction

Digital humans are an emerging technology that has been steadily advancing over time. Recent developments in artificial intelligence have significantly accelerated this progress, greatly expanding both the applicability and potential roles of digital humans. Unlike traditional virtual assistants, digital humans are distinguished by the presence of a human-like avatar, enabling more natural and expressive modes of interaction. This visual embodiment allows for the simulation of human communication cues such as eye contact, facial expressions, and emotional responses—features that can facilitate user interaction by allowing people to rely on familiar human-to-human communication strategies, reducing the need for prior system learning.

However, the inclusion of a face and other human-like traits also introduces new challenges. One prominent issue is the uncanny valley phenomenon, in which avatars that closely resemble real humans may evoke feelings of discomfort or eeriness. Throughout the technological development of digital humans, such challenges have posed barriers to acceptance and usability. Nevertheless, continued improvements in visual fidelity, animation, and behavioral realism may have brought the technology closer to a "sweet spot," where the benefits of lifelike interaction outweigh the perceptual drawbacks. In such a state, digital humans may effectively support roles traditionally performed by humans, such as receptionists or information kiosk attendants.

NTT Data, a company offering custom digital human solutions, currently operates two digital human installations in Aarhus, Denmark. This project is conducted in collaboration with NTT Data and focuses on improving the user experience of digital human receptionists.

Chapter 3 - Exploring the Problem Space

As public services experiment with AI-driven interfaces, the boundary between automation and human care becomes increasingly important to navigate. In collaboration with NTT DATA, this project explores what happens when a digital human replaces a traditional receptionist in a sensitive, emotionally charged environment. The following section outlines the practical and conceptual challenges involved in implementing Trothilde, a digital receptionist at Aarhus Jobcenter, and sets the stage for the fieldwork and design considerations that follow.

3.1 NTT DATAs Problem: Replacing the Receptionist without Losing the Human Touch

In our initial meeting with NTT Data, we were introduced to the central challenge of this case: a digital human has already been implemented at the entrance of the job center in Aarhus, intended to replace the role of a receptionist (see figure 3.1). She assists citizens with routine tasks such as check-in procedures and providing directions. However, despite her advanced capabilities, the way the digital human is currently used highlights a disconnect between what the technology can do and how people actually engage with it in practice.

Figure 3.1: The image presents an illustration of Trothilde as she appears in the job center environment

NTT Data's clear ambition is that it should not simply function as an interactive screen or self-service kiosk, but rather serve as a warm, welcoming presence that mirrors the empathy and attentiveness of

a real receptionist. Her presence is meant not only to guide citizens through the initial steps of their visit but also to ensure they feel seen, heard, and respected at a moment that can be emotionally taxing, especially for those facing unemployment or uncertainty about their future.

By taking over these routine and repetitive tasks, the digital human also plays a strategic role in relieving pressure on staff. This allows human employees to dedicate more time to providing in-depth support, especially in conversations that demand emotional sensitivity and professional nuance. In this way, she complements the job center's workforce by freeing up valuable time and enabling more meaningful, person-centered interactions between citizens and staff.

This ambition is not without complexity. The job center is often the first point of contact for citizens who are either newly unemployed or in a vulnerable life situation. As a recent cross-municipal citizen survey shows, younger job seekers (ages 16–29) are particularly likely to report feeling distrusted and emotionally unsupported during their job center visits. Many respondents express a strong preference for empathetic job counselors who acknowledge their life circumstances and communicate with warmth and clarity. (Alstrup, 2025)

NTT DATA has previously conducted observational fieldwork at the job center in Aarhus, where they examined how citizens interact with the digital human, Trothilde (Rebecca Hansen, 2025). Based on these observations, they reported several critical findings regarding user behavior and perception.

Currently, the digital human is often mistaken for a touch-screen kiosk. Positioned centrally in the room and speaking loudly, she paradoxically draws attention while also being perceived as difficult to engage with in a natural, conversational manner. Most users bypass her entirely, opting instead for the familiar tablet or seeking a human staff member. See the different paths in the Job center in figure 3.2.

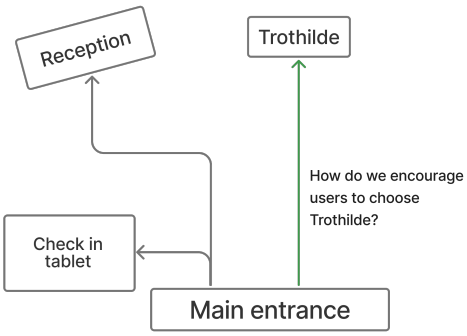

Figure 3.2: This illustrates the three pathways currently facilitated by the job center. The intended objective is to consolidate these into a single route through Trothilde.

At the core of this project is a conceptual tension: is the digital human intended to replace the tablet used for self-check-in, or is she meant to take on the role of a human receptionist? According to NTT Data, the intention is the latter. However, suppose citizens do not perceive the digital human as a social actor. In that case, the expected benefits, such as improved accessibility, emotional support, and increased staff efficiency, are unlikely to be achieved.

This gives rise to a dual design challenge. Firstly, it is necessary to ensure that citizens understand the digital human's purpose and feel motivated to interact with her. Secondly, the interface must be designed in a way that convincingly reflects the social and emotional qualities typically associated with a personal encounter at a reception desk.

3.2 Field study at job center Aarhus

Although NTT DATA had previously provided a detailed account of Trothilde's implementation at the Aarhus job center, the project group found it necessary to conduct an on-site investigation in order to observe user interactions in situ. To this end, an observational study was carried out on February 14th 2025, to gain firsthand insights into how the digital human is perceived and used by citizens.

The initial methodological plan was to conduct a rainbow usability sheet study in combination with short exit interviews. This mixed-methods approach was designed to capture both usability issues

and subjective user experiences, offering a comprehensive perspective on Trothilde's effectiveness. However, upon arriving at the job center, it quickly became clear that many of the visitors were either under time pressure or disinclined to participate in interviews, likely due to the sensitive nature of their visit.

As a result, the research design was pragmatically adjusted to focus exclusively on non-intrusive, passive observation, and the originally planned rainbow sheet method was abandoned due to the limited opportunity for structured usability feedback. Adopting a "fly on the wall" approach allowed the researchers to systematically document user behavior without interfering with the natural flow of interaction. Additionally, the researchers themselves engaged with Trothilde as end users in order to explore her functional capabilities and technical limitations. This revised approach proved valuable for generating empirical insights into both user behavior and system performance in a real-world setting.

Building on this adaptive approach, the following sections presents selected observations from the field study, organized into two categories: (1) citizen behavior observed at the job center and (2) reflections based on the research team's interaction with the digital human. (see observations in attachments "Jobcenter observations").

3.2.1 Observations of Citizen Behavior

The field study revealed several patterns in how citizens engaged with the physical space and the available check-in technologies. Notably, the digital human was not interacted with via speech during the observations.

In one instance, a young man initiated contact with Trothilde, via the touch screen, which triggered a chain reaction: others queued behind him and proceeded to check in using the same interface. Conversely, when one person opted for the traditional touch screen instead, the following citizen mirrored that behavior, suggesting a social imitation effect among users.

Users tended to prioritize approaching either the traditional check-in tablet or the human receptionist, as illustrated in figure 3.2. The human receptionist continued to play an important social role, greeting guests and offering help that extended beyond the job center's scope, for example, giving

directions to a municipal office across the street. Such behaviors highlight the subtle social and spatial awareness that Trothilde currently lacks.

3.2.2 Researchers' Reflections and Interaction

In direct use, Trothilde exhibited several limitations. When the researchers tried talking with Trothilde, she often interrupted the conversation when it wasn't her turn, and when confronted with questions outside her scope, she responded with "I don't understand," leading to the researchers repeating themselves to no avail, especially when questions were vague or complex. She pointed incorrectly when asked for directions and lacked feedback during system processes, appearing inactive. QR codes were provided for basic information, though these were not always accessible or useful.

Her loud voice compromised user privacy, and she failed to respond to simple greetings or cues like standing directly in front of her. Many users struggled with scanning their cards, leading to session resets. She also frequently interrupted the user mid-sentence, making elaboration difficult and disrupting the conversational flow. These limitations are visually summarized in figure 3.3, which outlines key barriers to interaction, including issues related to speech, interaction flow, graphical representation, and social behavior.

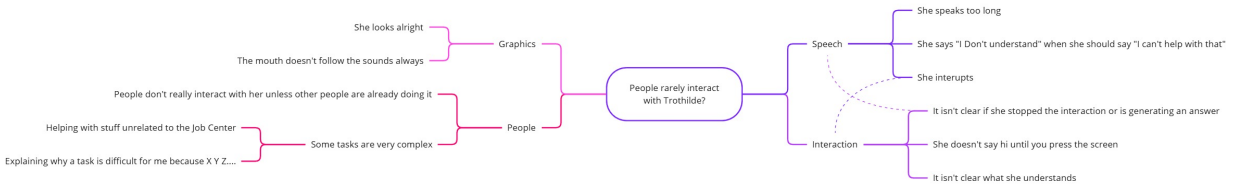

Figure 3.3: This visualization summarizes observational insights and researcher reflections regarding why citizens rarely engage with the digital human at the job center. The challenges are grouped into four categories: speech, interaction design, social behavior, and graphical representation. The map highlights issues such as unclear system feedback, limited conversational ability, lack of social cues, and hesitation to initiate interaction without observing others doing so.

3.2.3 Summary

Overall, the observational study provided valuable, situated insights into both the affordances and limitations of Trothilde in her current form. While not methodologically exhaustive, the study fulfilled its purpose as an exploratory tool for contextual understanding. It revealed clear patterns in user behavior, highlighted gaps between technical design and user expectations, and identified opportunities for more socially intuitive interaction design. The following figure has been developed to illustrate the key observations derived from the field study.

3.3 Interview at Krifa

Krifa was selected for an interview due to their distinct implementation of the digital human Sophia as the only point of contact in their reception area. Unlike the job center, where multiple parallel solutions coexist, Krifa fully restructured their lobby when Sophia was introduced, ensuring that all visitors are funneled through the digital receptionist. As a result, Sophia is used more consistently and visibly than Trothilde at the job center.

This difference in implementation context offers valuable insights into how organizational commitment and spatial design influence user interaction. To explore these dynamics further, we conducted a semi-structured interview with a staff member at Krifa, who shared practical experiences, challenges, and observations regarding the use of their digital receptionist, Sophia.

The interview revealed several recurring issues. A key challenge is visibility: many users overlook Sophia or fail to recognize her as an interactive point of contact. User behavior also appears to be shaped by habit—visitors often head straight for the touchscreen or follow what others do, which leads to the voice interaction method not being used. In terms of functionality, slow response times, limited system feedback, and occasional technical malfunctions contribute to user frustration and disengagement. While certain user groups, such as business clients and younger visitors, tend to respond positively to the technology, others remain hesitant and prefer traditional human interaction. Despite the automation of routine tasks, staff are still frequently required to step in and assist, especially when confusion arises.

These insights are based on a single staff member's perspective and may not reflect the full range of experiences across the organization. The complete interview is included in appendix A. Nonetheless, the account highlights several areas for improvement, including clearer signage, improved reliability, and more seamless integration of the digital human into the social fabric of the reception space.

3.4 Literature review

The information given by NTT and the observations from Aarhus job center and Krifa presents some fundamental problems with digital humans. This includes the physical surroundings, the digital human's sensitivity to sounds, conversation etiquette and interaction options.
Going forward it is relevant to study whether these problems are specific to NTT's digital humans or across platforms and if so, what research has been done to investigate these problems. Therefore, a systematic literature review has been conducted (all materials for the review can be seen in attachment "Literature review"). The literature review follows the guidelines provided by vom Brocke et al. (2009).
The literature review has been carried out over two rounds. The purpose of the first round was to get an overview of the research areas within the field of digital humans, as this is a new working area for the authors. The purpose of the second round of the literature review was to do a deep dive into one research area and uncover the most common problems within this area of digital humans.
In the context of this report, a digital human is defined as a system which contains an avatar and is able to process and produce speech. However, for some topics, it was deemed relevant to also include articles that concerns speech recognition software that does not have an avatar. The literature contains several synonyms for digital humans and the following terms has therefore also been used to cover the topic:

- Embodied conversational agent
- Conversational agent
- Intelligent agent
- Intelligent virtual agent
- Embodied virtual assistant
- Virtual conversational agents

- Artificial-Social-Agent

The following databases were used during the literature review, as they contain journals regarding human-computer interaction:

- APA Psychnet
- Primo
- Primo AI
- Scopus AI
- Google Scholar

The search was limited to journals, conference papers and books published after 2010. The limit was set to 2010 since this marks the point for when AI-driven voice assistants such as Google, Alexa, Siri were released (Network, 2024).

3.4.1 Literature Review, Round 1

The search terms that was applied during the first round of the literature review are displayed in figure 3.4 below.

Search Term	Year	Criteria	Peer review	Search engine	Results	Rough sorting	Relevant
"Digital human" AND interact*	2010-2025	Title Abstract	☐	APA Psychnet	109	3	1
"digital human" AND interaction NOT Sex NOT Crime NOT Violence	2020-2025	Title Abstract	☐	APA Psychnet	71	14	5
Embodied conversational agents (ECA) and human-computer interaction	2010-2025	Title Abstract	☑	APA Psychnet	8	1	0
embodied digital agent	2010-2025	Title Abstract	☑	APA Psychnet	31	12	5
Virtual conversational assistants (VCA)	2010-2025	Title Abstract	☑	APA PsycNet	1	1	1
AI AND Anthropomorphism AND Peer review	2020-2025	Includes AI with face	☑	APA Psychnet	115	47	15
"Digital human" AND Anthropomorphism	2010-2025	Title Abstract	☑	APA Psychnet	1	0	0
"Digital human" AND UX	2010-2025	-	☑	APA Psychnet	0	0	0
"Digital human" AND Empathy	2010-2025	Title Abstract	☑	APA Psychnet	5	1	1
"Digital human" AND Social	2010-2025	Title Abstract	☑	APA Psychnet	52	3	2
prosody AND digital humans	2010-2025	Title Abstract	☐	APA PsycNet	12	2	0
Prosody AND embodied conversational agents	2010-2025	Title Abstract	☐	APA PsycNet	1	1	1
Prosody AND virtual avatars	2010-2025	-	☐	APA PsycNet	0	0	0
Prosody AND synthetic speech	2020-2025	Title Abstract	☐	APA PsycNet	13	4	0

Figure 3.4: This figure shows the search terms that has been applied during the first round of the literature review.

The search terms covers interaction with digital humans, anthropomorphism, UX, empathy, social, prosody, and searches of synonyms for digital humans.
These searches provided 31 relevant articles, 29 without duplicates. References to these articles are listed in table 3.1 below.

Anthropomorphism
(Abel et al., 2020), (Ahn et al., 2021), (Bailenson & Yee, 2005), (Blut et al., 2021), (Coghlan, 2024), (Fakhimi et al., 2023), (Geiselmann et al., 2023), (Lee et al., 2010), (Martin & Jenkins, 2024), (Sharma & Vemuri, 2022), (Ter Stal et al., 2020), (Watson et al., 2024), (Xavier et al., 2023), (Y. (Zhang et al., 2024), (Z. Zhang et al., 2025)
Prosody
(Ehret et al., 2021), (Roso et al., 2024)
Digital humans in general
(Abel et al., 2020), (Erengin et al., 2025), (Geiselmann et al., 2023), (Loveys et al., 2022), (Sharma & Vemuri, 2022), (Potdevin et al., 2021),(Guadagno et al., 2011),(Sung et al., 2023)
Uncanny valley
(Diel et al., 2022), (Stein et al., 2020)
UX
(Guadagno et al., 2011), (Bae et al., 2025)

Table 3.1: This table shows all the sources that was found and marked as relevant in round 1 of the literature review

Neither backward or forward search were applied and the articles were not analyzed. Instead, the articles were skimmed and used to map the most prominent research topics within the field of Digital humans. As can be seen in figure 3.5, the topics include appearance, face, entertainment, emotions, gestures, conversation, and speech.

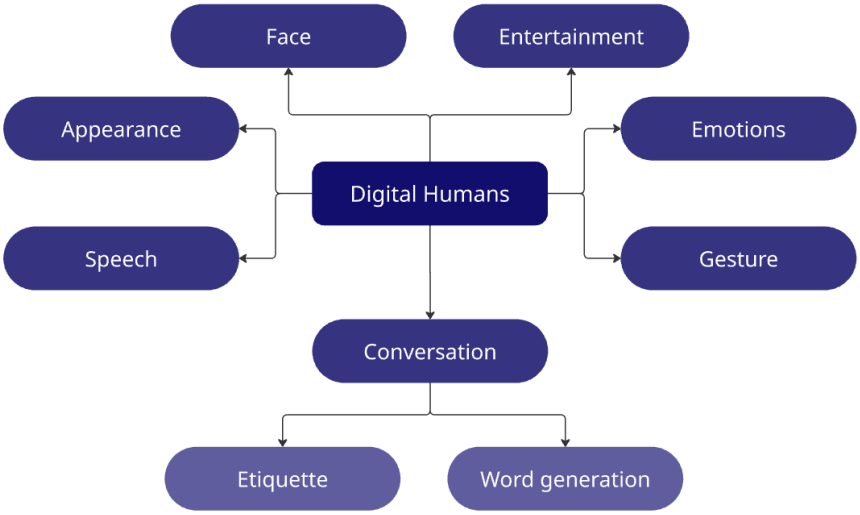

Figure 3.5: This figure shows a map of the most prominent research topics within the field of digital humans.

The map in figure 3.5 was used to narrow down the focus of the report to a topic that is both interesting, possible to research and relevant to the the problem areas that has been identified with NTT's digital humans. By now, we already know from NTT that we will not be able to change the graphical elements of NTT's digital human. This means, that the topics appearance, face, and gesture are excluded. Of the remaining topics, conversation was the most interesting topics as it feeds into the problems mentioned in section 3.2 with NTT's digital humans.

3.4.2 Literature Review, Round 2

In the second round of the literature review, only the articles about conversation related aspects were included. Additionally, eight new searches were made with APA Psychnet, Google Scholar, Scopus AI, and Primo AI. These searches provided 13 relevant articles. The search terms are listed in figure 3.6 and include conversation, AI latency, initiating conversations, emotional responses, embodied conversational agents vs. humans, and anthropomorphism.

Search Term	Year	Criteria	Peer review	Search engine	Results	Rough sorting	Relevant
AI AND Response time AND Conversation	2010-2025	Title Abstract	☐	APA PsycNet	6	0	0
AI latency AND conversation	2010-2025	-	☐	APA PsycNet	0	0	0
AI latency AND conversation	2010-2025	Title Abstract	☐	Google Scholar	22500	3	1
"I would like to view articles about digital humans, embodied conversational agents or virtual Conversational Agents and how people initiate conversation with them"	2010-2025	Title Abstract Description	☐	Scopus AI	11	7	4
Hvilke resultater kan man finde ved samtale mellem emodied conversetional agents og mennesker		Title Abstract Description	☐	Primo AI	5	4	3
What are the emotional responses of humans interacting with embodied conversational agents?		Title Abstract Description	☐	Primo AI	5	3	2
Er der modstridende resultater for samtale mellem embodied conversational agents og mennesker		Title Abstract Description	☐	Primo AI	5	3	3
antropomforsime i forhold til samtalen med embodied conversational agents		Title Abstract Description	☐	Primo AI	5	1	0

Figure 3.6: This figure shows the search terms that has been applied during the second round of the literature review. Note, that for the search on Google Scholar, only the first two pages were included.

These articles were read by the authors, and for each article snowballing was applied. Meaning, that the articles bibliography was skimmed to find other relevant articles. Snowballing provided 19 additional articles. All articles found in round 2 are listed with their reference in table 3.2 below.

Communication style
(Krämer, 2008), (Van Pinxteren et al., 2023), (Van Pinxteren et al., 2020)
Interaction
(Smith et al., 2010), (Foster, 2007)
Perception and accept of agents
(Pickard et al., 2014), (Chateau et al., 2005), (Biancardi et al., 2021), (Loveys, Sebaratnam, et al., 2020)
Agents in professional roles
(Sharma & Vemuri, 2022), (Čereković et al., 2009)
Development
(Dirk Heylen et al., 2009), (Esposito & Jain, 2016)
Snowball
(Akdim et al., 2021), (Baron, 2015), (Becker et al., 2007), (Benitez et al., 2017), (Ben Mimoun & Poncin, 2015), (Bergmann et al., 2012), (Bickmore & Cassell, 2005), (Cerekovic et al., 2017), (Chattaraman et al., 2019), (Elliot, 1996), (Gratch et al., 2007), (Groom et al., 2009), (Kopp et al., 2005), (Kramer, 2005), (Loveys, Sagar, & Broadbent, 2020), (Oertel et al., 2020), (Robinson et al., 2008), (Taylor & Baker, 1994), (Wagner et al., 2019)

Table 3.2: This table shows all the sources that was found and marked as relevant in round 2 of the literature review

All dependent and independent variables in the articles were noted in an Excel document along with their effects. This was done to create an overview of what elements affect digital humans and how they affect. The full list of variables can be found in appendix B or attachment "Literature review". A summary of the variables is made in table 3.3 below.

Dependent variable	Independent variable	Effects
User behavior, including attitude and communication	Human-like appearance	More human-like appearance leads to conversations with natural speech (Kramer, 2005)
Anthropomorphism	Human-like features, demographics	More human features equals more anthropomorphism (Blut et al., 2021). Elderly, children and women anthropomorphize more (Blut et al., 2021).
Affect, likability, satisfaction, rapport, perceived usefulness, presence, trust, safety, and perceived intelligence	Anthropomorphism	More anthropomorphism increases the perception of the dependent variables (Blut et al., 2021)
Persuasion	Anthropomorphism, user personality, perceived capabilities, value of task, and type of message	Anthropomorphism increases persuasion (Blut et al., 2021). User personality affects persuasion (Watson et al., 2024). Simpler, desirability focused or utilitarian tasks makes users more prone to persuasion (Watson et al., 2024)
Rapport	Communication style (social vs. task, reciprocity principle, hedonic vs. utilitarian)	Social and adaptive communication style increases rapport (Van Pinxteren et al., 2023). So does reciprocation and hedonic communication (Chattaraman et al., 2019; Van Pinxteren et al., 2023). Rapport increases satisfaction and intention of use (Van Pinxteren et al., 2023)

Trust	Small talk, appearance and personality	Introverts trust task-oriented agents more. Extroverts trust agents that small talk more (Bickmore & Cassell, 2005). Digital humans rates higher on trust than non-embodied agents (Ben Mimoun & Poncin, 2015; Blut et al., 2021).
Uncanny valley	Voice distortion, motion manipulation, visual-auditory mismatch	All three independent variables has a moderate effect on controlling the response to uncanny valley (Diel et al., 2022)
Engagement, trust and cognitive load	Task based vs. social based dialog	Users trust and engage with digital humans more when they match the users conversation style and degree of extroversion/introversion (Bickmore & Cassell, 2005). Social based dialog creates less cognitive load (Chattaraman et al., 2019).
Linguistic behavior	Perceived power of the avatar Perceived likability of the avatar	Female avatars has a higher perceived likability and lower perceived power. When an agent is perceived as powerful, participants used less expressive language, more hesitations, and they revealed less about themselves. (Pickard et al., 2014)

Other		Agents saying "I" are perceived as more empathetic (Van Pinxteren et al., 2020). Personalization improves interactions (Y. (Zhang et al., 2024). Apologies make computers seem more sensitive (Van Pinxteren et al., 2020). Higher psychological closeness facilitates better responses when the agent focuses on a products secondary features (Ahn et al., 2021). Intelligent agents reduce search time and costs (Ben Mimoun & Poncin, 2015). Friendly language increases warmth, and text-to-speech has better flow than text-only communication (Ben Mimoun & Poncin, 2015).

Table 3.3: This table shows the independent variables, the dependent variables and their effect on one another that has been identified through the literature review.

3.4.3 Conclusion on Literature Review

Across the reviewed literature, a clear methodological pattern emerges: the majority of studies investigating user interaction with conversational agents and embodied conversational agents (ECAs) rely on quantitative, experimental designs, often involving manipulated variables such as communication style (social vs. task-oriented), degree of personalization, anthropomorphism, or voice characteristics (e.g., human vs. synthetic).

Several studies employ survey-based approaches to measure user perceptions, while others contribute with systematic literature reviews or qualitative analyses to provide contextual depth. Common dependent variables include engagement, rapport, user satisfaction, perceived usefulness, trust, presence, anthropomorphism, uncanny valley, and behavioral intention, often measured using 5- or 7-point Likert scales.

3.5 From Observations to Design Inquiry: Identifying Key Areas of Interest

Building on the insights gathered from both the literature review and on-site observations at the job center, several recurring themes have emerged that highlight critical friction points in interactions with digital humans.

Drawing on this synthesis of empirical and theoretical perspectives, we have identified three interrelated areas that merit closer examination. These areas will serve as a foundation for the subsequent sections, in which we explore how targeted design interventions might help address the challenges observed, while also considering how these efforts are informed or potentially complicated by existing research.

1. Physical context and conversational space
2. Task-based versus social-based conversation modes
3. Initiating interaction

The first area of interest concerns the physical context in which the digital human is embedded. This focus emerged directly from our field observations at the job center, where it became evident that users did not initiate spoken interaction with Trothilde. In the majority of observed cases, individuals either bypassed her entirely or engaged solely through the touchscreen interface. No users attempted to speak with her, and even in instances where she responded verbally, users did not treat her as a conversational partner. This suggests that the system fails to afford voice-based interaction in a meaningful or perceivable way.

In design theory, affordances and signifiers play a crucial role in shaping how users interpret the purpose and functionality of an object or interface (Norman, 2002). While Trothilde is technically capable of audio-based dialogue, this potential is not communicated clearly through her physical presence. Instead, the interface is dominated by touch-oriented cues such as the placement of the touchscreen and on-screen instructions which overshadow any subtle signifiers of spoken interaction. As a result, users do not perceive Trothilde as a system that invites conversation, but rather as a passive kiosk or information terminal.

This topic would include an investigation of how the spatial and sensory environment can be redesigned to promote social cues, reduce ambiguity, and increase users' willingness to engage vocally with the digital human.

The second area of interest is the adaptive conversational styles. This behavior is supported in making the interaction better with increases in rapport, trust and engagement. An implementation of this in Trothilde should enhance these variables. By distinguishing between task-oriented and socially-oriented modes of interaction, researchers have identified important differences in how users relate to and evaluate conversational agents (Bickmore & Cassell, 2005; Chattaraman et al., 2019). Task-oriented dialogue tends to focus on the efficient execution of predefined goals, while social-oriented interaction introduces relational cues such as small talk, empathy, and politeness strategies to simulate a more human-like presence. The choice of interaction style significantly shapes user engagement and satisfaction, especially in emotionally sensitive contexts such as job centers, where users may be navigating uncertainty or stress (Ben Mimoun & Poncin, 2015; Blut et al., 2021; Van Pinxteren et al., 2023).

Prior studies suggest that social-style agents can enhance users' perceptions of warmth and trust, particularly when the user lacks confidence in their task competence (Chattaraman et al., 2019). Conversely, users with high digital literacy or goal-oriented mindsets may prefer direct and efficient task-based interactions (Van Pinxteren et al., 2023). This highlights the need for adaptive systems that can modulate their communication style based on contextual and user-specific cues (Blut et al., 2021).

For our project, prototyping distinct conversation modes offers an opportunity to explore how interaction framing whether through verbal tone, conversational pacing, or user-initiated mode switching affects perceptions of empathy, responsiveness, and usefulness. As Bickmore and Cassell (2005) emphasize, socially intelligent agent behavior must be matched to user expectations and situational needs in order to foster rapport and engagement. Exploring this adaptive capacity is therefore a key step in designing digital humans that can support both the informational and emotional dimensions of user experience.

The third area of interest is agent-initiated conversation, which appears to be underexplored in the existing literature. Most commercial voice assistants currently available—such as Siri and

Alexa operate primarily in a passive mode, where interaction is initiated by the user through explicit verbal cues like “Hey Siri” or “Alexa.” In these systems, the burden of initiating conversation rests with the user, rather than the system. This stands in contrast to typical human receptionist–customer interactions, where the receptionist is generally expected to initiate engagement. Because this form of agent led initiation has not received substantial attention in current research on conversational agents, it presents a particularly compelling area for further investigation.

Recent experimental studies have begun to probe the design potential of more proactive agents. For example, Tabalba et al. (2024) conducted a study using ArticulatePro, comparing two versions of a voice assistant: one reactive and one proactive. I.e. the AI will ask follow up questions or answer slightly outside the scope. The proactive version not only resulted in faster task completion during a data analysis task but also led to higher user engagement. Users interacted for longer durations and in a more dialogic manner because the assistant actively offered follow up suggestions and next steps effectively leading the interaction when it served the task at hand. This highlights how proactive behavior can encourage sustained and meaningful dialogue.

The third area also overlaps conceptually with the first, as both focus on challenges related to initiating conversation. However, while the first area emphasizes traditional design principles to address this issue, the third focuses more specifically on the conversational roles and expectations typically found in human service interactions. Hopefully, the initiation of interaction will serve as a strong enough signifier that the affordance of speech interaction is present and users use it to have easier interactions with the digital humans.

Taken together, these three areas point to a broader challenge: how to design systems that support more intuitive and engaging conversational experiences with digital humans. While each area offers valuable insights, for example, the second area highlights the need for adaptive conversational styles, there is chosen to prioritize the third area, agent-initiated conversation, as the central focus of our design inquiry. This area stands out not only due to its relative lack of attention in the literature but also because of its strong connection to behaviors observed in our field study and its potential to fundamentally shift how users perceive and respond to digital human interactions.

By exploring how digital humans might adopt a more proactive role in initiating conversation, much like human receptionists or service personnel, we seek to understand whether such an approach could help reduce hesitation and encourage greater user engagement in socially situated contexts.

3.6 Problem statement

Interestingly, both the literature and our own fieldwork raise important questions about how context and initial impressions shape users' willingness to engage with digital humans. Why do so many passersby ignore Trothilde, even when she is prominently positioned and technically able to converse? It appears that the very first moments - the way the interaction begins, are decisive. Users seem to look for subtle but powerful social cues that signal: "This is a space where I can speak, and I will be heard." But what are those cues, and how do we design for them?

Building on this observation, the present study investigates how different approaches to initiating conversation, specifically whether the interaction is initiated by the user or by the digital receptionist, influence the level of user engagement. By focusing on the critical entry point to the interaction, we aim to better understand what cues matter and how they might be designed to foster more approachable and socially intuitive encounters with digital humans.

This leads us to the following problem statement:

> *What differences can be observed in user behavior and engagement when interactions with digital humans are initiated by different parties?*

Chapter 4 - ROSIE — Receptionist Optimized for Student Information & Engagement

Rosie is the digital human developed specifically for this study. While initial plans included using the existing avatars such as Trothilde and Sophia, these models lacked the flexibility required to support the targeted interaction styles identified during the literature review. Therefore, a custom solution was developed to ensure greater control and experimental flexibility.

This custom solution should live up to the following list of requirements:

1. It should be possible to use different dialogue styles or make a prompt to perform an interaction in a certain manner.
2. It should be based on a Large Language model, to support flexible dialogue management.
3. It should be able to listen to audio from a microphone.
4. It should be able to return audio to a speaker.
5. It should have an animated face.

Regarding the face, it is important to consider the fidelity of the face as to not create the uncanny valley response. This means that it should either be very high fidelity moving beyond the valley or a sufficiently lower fidelity as to never get into the valley Ferdinand Engländer, 2014."

For this purpose the Nvidia audio2face-3d model fulfilled the requirements. It uses the flagship 3d-model that Nvidia uses to showcase how advanced their GPU's and artificial intelligence has become and is very high fidelity. Additionally it is very simple to use via the Omniverse client or via their docker containers called "Nvidia Inference Microservices" NIMS. The docker container however presented significant implementation challenges and as such the omniverse client was chosen. This client requires that the audio is sent via the gRPC protocol. An example of how to use the gRPC protocol is present in the code for the audio2face-3d program and this example has been modified to fit the project.

The NIMS also include LLM's but as it wasn't possible to get them to run properly, most probably due to inexperience with setting up a linux environment, other alternatives had to be considered. The OpenAI API seemed a fitting choice as their AI "ChatGPT" is very popular and scores well as a LLM.

Investigations into OpenAIs API presented the OpenAI Agents pipeline which can be used to send and receive audio files. The processing all takes place on OpenAI's servers and it all happens relatively fast 5-10 seconds. In addition, OpenAI has excellent documentation[1] with examples for our use case (OpenAI, 2025b).

An alternative to the Agents pipeline would be to chain together multiple AI's. One for transcription, then inference and then text to speech. This could give more flexibility and the ability to fix issues down the road but most probably would also introduce more issues. Therefore the agents pipeline has been chosen for the project. Using a cheap model such as gpt-4o-mini means that the processing for the entire project can probably be done for less than 20€. When development for the first iteration was over, about 1€ had been spent.

4.1 Programming Rosie

Rosie flow is in simple terms:

1. Record audio
2. Send audio to OpenAI
   (a) Speech To Text
   (b) Inference
   (c) Text To Speech
3. Receive audio from OpenAI in chunks and combine them
4. Send audio to NVIDIA Audio2face-3d via gRPC
5. Render face animations and play sound via Audio2face-3d
6. Repeat

Rosie is programmed in Python based on an example from OpenAI[2] and the Omniverse Audio2face-3d client example from the Omniverse Audio2face-3d application (in the Audio2face-3d package folders). (OpenAI, 2025b)

The sound handling is written in collaboration with the AI tool Github Co-pilot. This tool allows for easily writing correct syntax by instructing the AI to handle specific use cases such as storing

[1] https://openai.github.io/openai-agents-python/voice/quickstart/
[2] https://openai.github.io/openai-agents-python/voice/quickstart/

audio or manipulating it. The initial code is reviewed by the authors and adapted as needed. Some areas of the code has been written entirely by hand as the AI code was not good enough, most of the code has been modified and some hasn't been modified at all.

A showcase of Rosie can be seen on YouTube[3] (Jeppe Givskud, 2025a).

First the Agent is defined and loaded into the pipeline which is then run. The pipeline runs until the program is stopped manually. For the first interaction a simple 2 seconds of silence is sent to OpenAI's API which then responds with an audio clip of Rosie saying hello. When this is received the audio is bundled into a variable and the program now awaits the signal from the facilitator to send this audio to omniverse audio2face-3d to be played.

A GUI has been programmed for this purpose as well as for logging purposes. The GUI can be seen on figure 4.1. When the facilitator presses "Stop recording" the audio is formatted and sent to the OpenAI API via the pipeline.

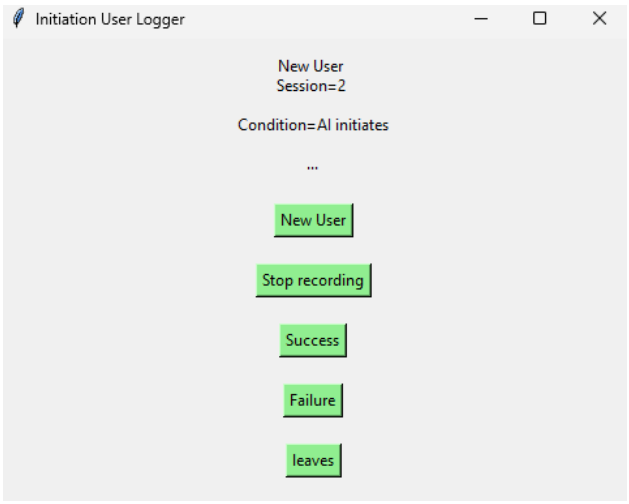

Figure 4.1: This figure shows the GUI used to record the outcome of the tests as well as control the conditions

When the signal is received by audio2face-3d the 3D model starts "speaking" and the audio and facial animation match in a coherent manner. After the audio is done the face freezes in the last position received. This is unfortunate as the last position can be during a blink or another weird facial posture. Therefor a very short audio signal is sent to hopefully reset the animation which works most of the time.

[3] https://www.youtube.com/watch?v=gKidbRIFDds

A summary of the functions in Rosie can be seen below in figure 4.2. This shows how many different functions interact in order to record sound, generate inference, generate audio and show a face.

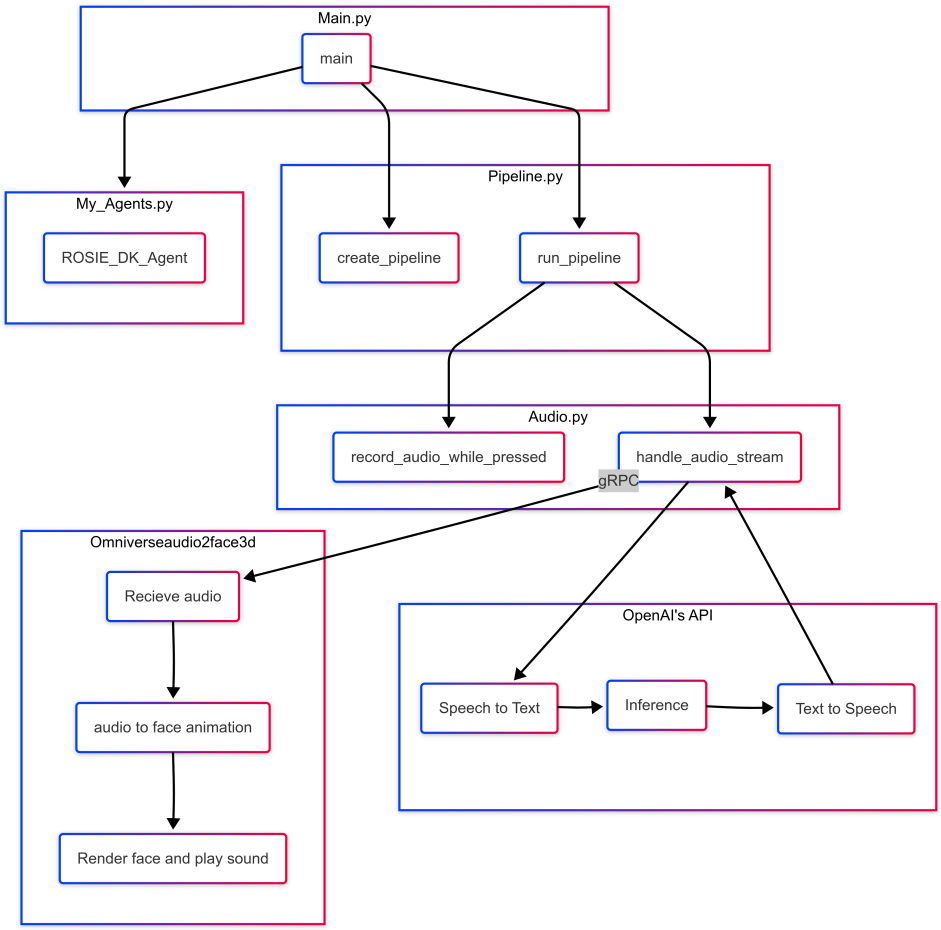

Figure 4.2: This figure shows a flow chart for how the functions and programs behind Rosie interact with each other.

4.2 Instructions for the AI

The OpenAI API allows for the creation of "Agents" via a simple text string. Agents are containerized AI models with different instructions and different functional access i.e. weather data. This is not to allow AI's to handover control making for task optimized AI interaction. In this research there has only been developed one AI agent -Rosie. The instructions for Rosie can be found in appendix D. As the agent will be interacting with danish students the instructions are in danish.

OpenAI suggests writing a more fulfilling description than "be polite" which handles use cases such as users saying no or which questions to ask and in what order. The personality in appendix D was written in collaboration with OpenAI's Voice Agent Metaprompt GPT (OpenAI, 2025a).

During development, it became apparent that Rosie lacked visibility into her system status, making it difficult to understand whether she was actively listening or processing input. To address this, a simple command-line interface (CLI)–based status indicator was implemented. It displays one of two words in ASCII art at the bottom of the screen: "LYTTER" (Danish for "Listening") when the microphone is actively recording, and "TÆNKER" ("Thinking") when the recorded audio is being processed by OpenAI's speech and language models.

Chapter 5 - Initiation of conversation

Before designing the experiment to answer the problem statement it is beneficial to explore exactly what it entails.

The problem statement concerns the observation of changes in user behavior and engagement. User behavior is a broad term to observe and this is deliberate as topic of conversation initiation isn't narrowed down to any specific constructs yet.

5.1 Hypothesis

This research explores whether AI-initiated conversation influences both the likelihood of interaction and the user's subjective experience. Specifically, does the AI initiating a conversation make users more likely to engage, and does it shape how the interaction is perceived?

There's reason to believe it might. In typical human-service worker interactions, it's common for the service provider to initiate contact, offering assistance. If AI mirrors this behavior, it may create a more natural and comfortable experience. Additionally, since digital human technology is still emerging, users may be unsure of its capabilities. Proactive initiation by the AI could signal that voice interaction is not only possible but smooth and intuitive.

As such the following two sets of hypothesis will be tested:

Likelihood to engage in conversation:

- $\mathbf{H0_1}$ – There is no difference in the rate of conversation initiation, regardless of whether it is initiated by the AI or by the user.
- $\mathbf{H_1}$ – There is a difference in the rate of conversation initiation between the two conditions.

Subjective experience of conversation:

- $\mathbf{H0_2}$ – There is no difference in the subjective experience of the conversation, regardless of whether it is initiated by the AI or by the user.
- $\mathbf{H_2}$ – There is a difference in the subjective experience of the conversation, between the two conditions.

5.2 Experimental Setup

This section presents the experimental foundation for investigating how to encourage conversations with a digital human (Rosie) in public spaces. Due to the unpredictable nature of conducting field experiments in semi-controlled environments, the setup evolved over time through iterative adaptation. This section includes a material overview, the core experimental design, and reflections that pave the way for the methodological approach detailed in section 5.3.

5.2.1 Materials

The experiment required the following key components:

- **Digital human agent, Rosie** (see section 4): A conversational AI embodied as a human-like figure, displayed on a vertical screen and capable of speech-based dialogue. The omniverse audio2face-3d configuration file can be found in Attachments "Rosie - Source code - IT3.usd" the computer used could render this face with 30-50 FPS.
- **Physical setup**: Including a vertical screen (Model: AOC 27G2U/BK 1920x1080), room dividers (to manage visibility) and environmental markers (e.g., floor tape or visual boundaries).
- **Audio and video recording tools** for transcription and behavioral analysis: Microphone (Model: Blue snowball), Speaker (Model: fostex model 6301B2AV), webcam for observation (Model: W88 S 1080P)
- **Survey tools**: Survey exact with a digital questionnaires based on TAM and TPI (see appendix C for further discussion, and attachment "Questionnaire" for the full questionnaire).
- **Visual aids**: Posters and QR codes to help explain the experiment without requiring direct facilitator intervention (see attachment "Posters for experiment".)

5.2.2 Experimental Design and Setup

While much of the existing research has been conducted in laboratory settings or highly scripted environments, this study aims to bridge the gap between controlled conditions and naturalistic use contexts. By designing a scenario that invites spontaneous interaction, it is possible to observe real-world behavior, rather than rely solely on imagined or retrospective self-report.

The experiment is an observational study that takes place in a public space to reflect the context of NTT's digital humans. To assess these conditions in naturalistic settings, four successive setups were implemented across two locations: the Atrium at Electronic systems at Aalborg University and Aalborg University Library Kroghstræde. Here, participants encounter a digital receptionist in an open, unsupervised space.

These two locations were chosen because they receive a lot of traffic from students. At the building for Electronic Systems, the experiment were set up in three different locations during the experimental sessions. These locations can be seen in figure 5.1 that also contains an assessment of the usual walking paths in the atrium, which is based on observations from the experiment.

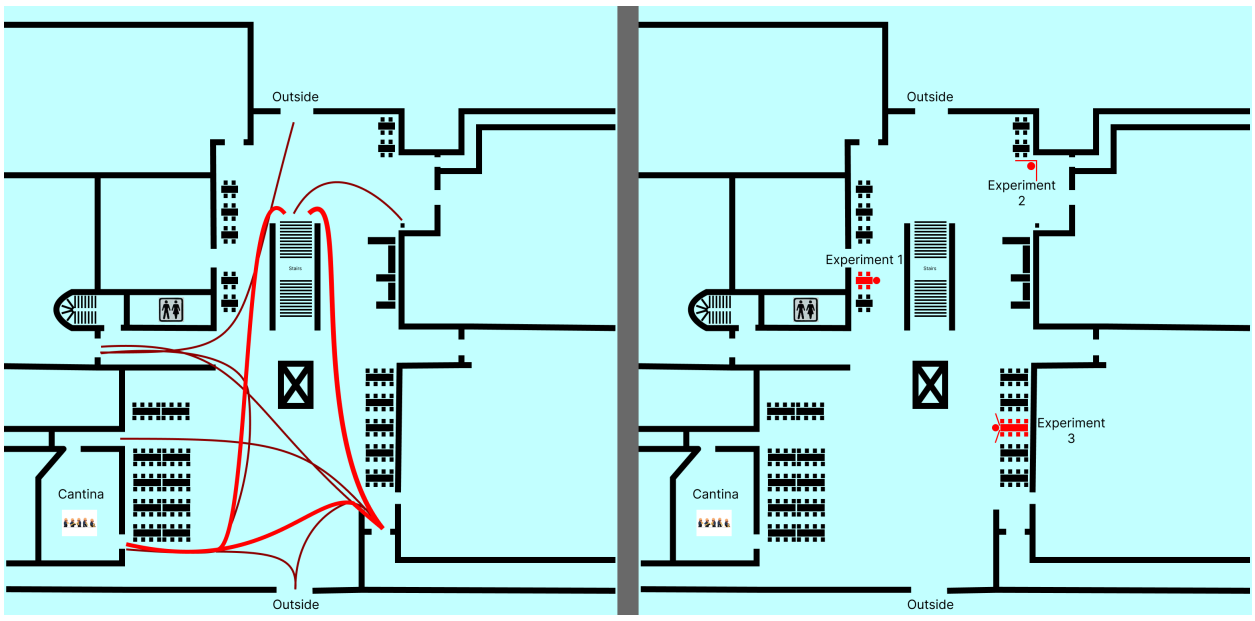

Figure 5.1: This figure shows a subjective assessment of how people usually walk in the atrium. It is drawn based on the observations from this experiment. The three experimental locations of the experiment is mapped

The experiment aimed to compare user engagement and interactions across two conditions:

- *Rosie-initiated interaction, where Rosie greeted passersby.*
- *User-initiated interaction, where the AI remained passive until the participant chose to speak.*

However, in order to quantify wether a participant should count as a success (interacted with Rosie) or failure (didn't interact with Rosie), it is important to define when a user has been "able to interact" with Rosie and should be included as a participant. In order to do so, an area around Rosie should be agreed upon, such that users who walk in this area are part of the study. To keep a balance between including enough users and having users close enough to Rosie, the area extended 2 meters in front of Rosie and 1 meter on each side of her. In this area, three posters saying "Talk to me" are placed around Rosie to attract participants.
When users set foot in the area, they would either be in the user-initiated condition and Rosie would do nothing, or they would be in the Rosie-initiated condition and Rosie would greet them. The condition changed for each participant.
If users walked out of the area without saying anything, they would be marked as a failure. If they stayed to talk, they would be marked as a success.
In a successful interaction, Rosie would go through the script as shown in appendix E, which can be summarized to the points below. The conditions are completely the same, except that the user start the conversation by saying hello in the user-initiated condition:

- Greeting
- Consent
- Project topic
- Project Status
- Challenges
- Help
- Goodbye

After ended conversation, the participants are asked to do an exit interview and questionnaire.
As the experiment were performed over several sessions, several iterations was done of the setup. Each iteration incorporated insights and adjustments based on observed user behavior, environmental affordances, and technical challenges.

5.3 Method

This section outlines the method used in the current experiment, which evolved over time and is therefore presented in four distinct "setups." Due to the unpredictable nature of conducting the experiment in a public setting, several methodological issues emerged gradually. Rather than adhering rigidly to a flawed initial design, the facilitators chose to adapt the method. This adaptive approach was deemed preferable, as accepting known methodological errors would have compromised the validity of the results to a greater extent than refining the method.

5.3.1 Data Collection Overview

Data was collected through:

- Exit questionnaires, including both scaled items (TAM, TPI) and background questions.
- Short exit interviews, capturing first impressions and reflections.
- Real-time observational notes, taken by facilitators during the experiment.
- Quantitative data: Time of interaction, Condition, Result of interaction
- Audio data: Audio is recorded and transcribed via Whisper (OpenAI, 2025c). This data is then analyzed for "Average sentence length", "Total conversation time", "Total turns".

This mixed-method approach allowed us to triangulate subjective experience, behavioral patterns, and attitudinal outcomes.

5.3.2 Qualitative data

Participants were asked a series of background and open-ended questions to explore their engagement and reflections on the interaction. Observational notes were also taken during each session.

The following questions were included in the survey to gain an initial understanding of participants' familiarity with and usage patterns of AI technologies. These background items served to contextualize the subsequent evaluation of the interaction with the digital agent.

- How often do you interact with AI-based tools or systems?
- Which AI services or tools have you used in the past 3 months?
- What do you use AI for when you engage with it?
- How do you typically interact with AI?
  - When do you prefer speaking, and when do you prefer writing?

The following open-ended questions were asked as part of the exit interview to explore participants' subjective impressions of the interaction with the AI agent Rosie. The aim was to capture spontaneous reflections and understand the factors that shaped their engagement.

- What was your first impression of Rosie?
- What made you stop and talk to her?
- Overall, how do you think the conversation went?

All data was analyzed using a reflexive thematic analysis inspired by Braun and Clarke's six-phase framework (Braun & Clarke, 2006). The analysis focused on identifying recurring patterns in user experience across the two experimental conditions: one where the user initiated the conversation, and one where Rosie initiated it. Quotes were coded inductively and grouped into overarching themes and subthemes.

The quotes used to showcase themes are translated from Danish to English by the authors.

To complement this qualitative insight, a structured questionnaire designed to systematically assess users' subjective experience was also included. Whereas the interviews provided open-ended reflections and behavioral observations, the questionnaire allows quantifying engagement and intention of use, by using validated measurement scale.

5.3.2.1 Measuring Engagement and Intention of Use

The reviewed literature identifies key design parameters that influence how users experience and interact with digital humans. Parameters such as communication style, personalization, embodiment, and agent personality have repeatedly been shown to impact outcomes like engagement, rapport, trust, and intention to use. These studies largely rely on quantitative methods, including Likert-scale surveys and experimental manipulations of agent characteristics.

Building on these findings, our study applies a combined quantitative and qualitative approach to assess how participants experienced their interaction with the AI agent. Specifically, we draw on the Technology Acceptance Model (TAM) to assess intention to use, informed by the operationalizations presented in Blut et al. (2021). TAM remains one of the most widely applied models for understanding technology adoption, linking constructs such as perceived usefulness and ease of use to behavioral intent.

In addition, to assess user engagement and perceived presence, we incorporate selected subscales from the Temple Presence Inventory (TPI) (Esposito & Jain, 2016, p.93-113). TPI is a validated multidimensional instrument designed to capture different forms of presence (e.g., social, physical, self) in human-technology interactions. This enables us to move beyond surface-level interaction and explore how socially embedded and immersive the experience with the digital agent feels.

The exit questionnaire included items from both TAM and TPI. For TPI, we specifically focused on the engagement subscale, which captures psychological involvement through dimensions such as mental immersion, sensory engagement, and perceived realism. These aspects are particularly relevant when assessing how absorbing or natural an AI-mediated conversation feels to the user—and whether this is influenced by who initiates the interaction.

In parallel, TAM items were included based on the original model developed by Davis (1989) and further expanded in other studies (Davis, 1989; Park, 2009; Venkatesh & Davis, 2000). In this framework, Perceived Usefulness (PU) and Perceived Ease of Use (PEU) are assumed to jointly shape the user's intention to use a system. While intention isn't measured directly, the use of the model follows this theoretical logic and serves as a proxy for evaluating user acceptance.

As the original TPI and TAM scales are written in English, a systematic translation process was carried out to adapt the items for Danish-speaking participants. This procedure followed core principles of cross-cultural scale adaptation, including forward translation, synthesis, back-translation, and review by external experts. The full theoretical background, item set, and translation protocol are documented in appendix C. The translations and validation can be found in attachment "Exit interview - Translation of questions".

5.3.3 Quantitative data

The quantitative data collected in this study is described in figure 5.2.

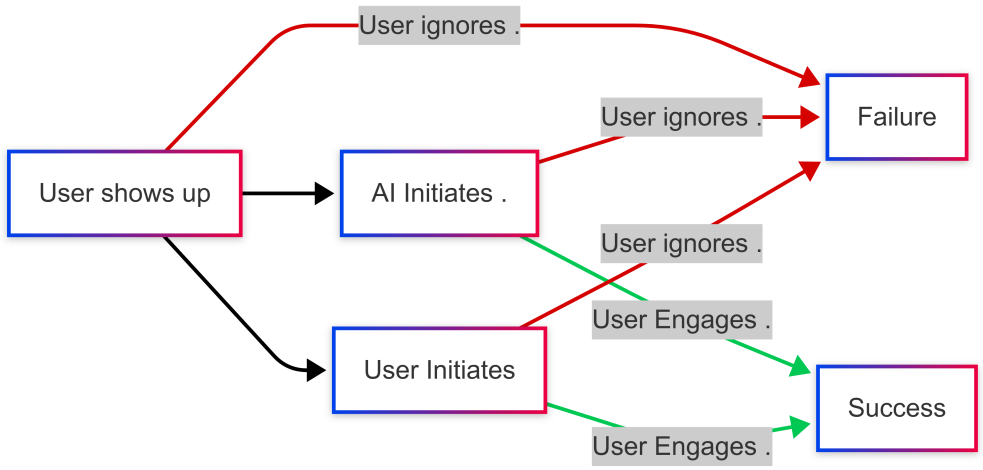

Figure 5.2: This figure shows how to quantify the different outcomes of the experiment in order to answer H_1.

The GUI presented in section 4 is used to classify each interaction with Rosie as either a success—indicating that the participant engaged in conversation or a failure, where the participant ignored or otherwise did not interact with Rosie.

In addition, audio from both Rosie and each participant is recorded and transcribed. The transcriptions will be analyzed for sentence length, overall conversation length, and the total number of conversational turns. Perhaps, if users have longer conversations and each sentence is longer this could indicate that the quality of interaction is better. This is not a focus of the study but the data is collected and analyzed for exploratory purposes.

5.3.4 Setup 1

Setup 1 was used for two session and was placed at the spot that is marked with *Experiment 1* on figure 5.1. The location can also be seen in figure 5.3 where the area that defines when people are able to interact is marked with red.

Figure 5.3: When users set foot on the red area Rosie will either do nothing or she will engage the user with the normal greeting. If users walk out of the area without saying anything it will be marked as a failure. If they stay and talk it will be marked as a success. If they say hi but then change their mind and leave it will be marked as "success - user leaves".

5.3.4.1 Results - Setup 1

Across the two sessions where setup 1 was used, a total of 60 participants was included. An overview of all trials can be seen below on figure 5.4.

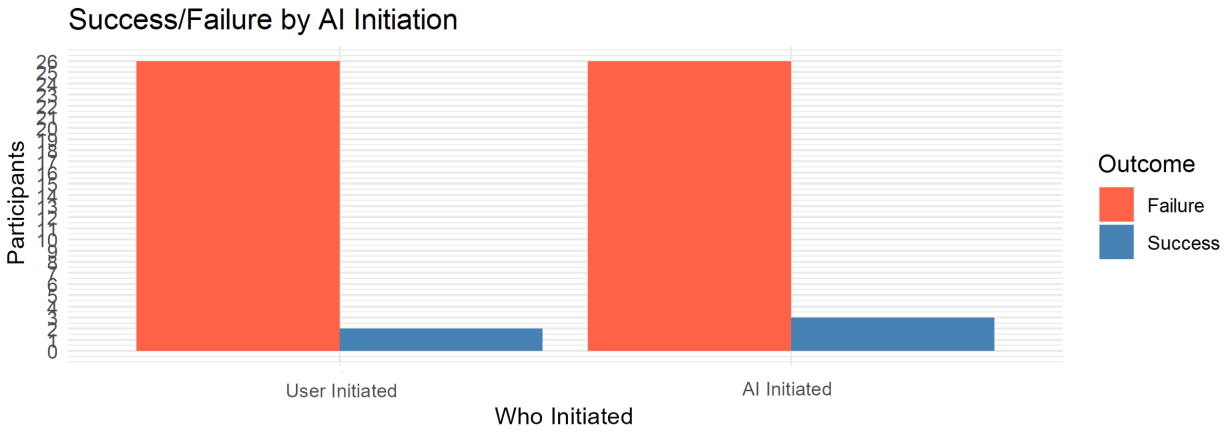

Figure 5.4: This figure shows the distribution of failures to success between the two conditions AI initiates and user initiates

5.3.4.2 Discussion - Setup 1

As almost no participants engaged with the AI and the difference between the two conditions are 1 participant the experiment would need to be modified for more engagement.

When users walk past without facing Rosie they might not see her and as such cannot "ignore" her. However these were still marked as failures.

As the experiment takes place in a public space where people can walk past several times, some participants might not be have been recognized by the facilitators as reoccurring and is therefore registered more than once. If the facilitators recognize a reoccurring participant, they won't be marked again.

When multiple users are present in the area during an instance where Rosie initiates a conversation, the interaction is counted as a single participant, regardless of how many individuals are nearby. This approach slightly affects the data, as users who do not respond to the initiation are not recorded as failed interactions.

The experiment shows that users are not very likely to engage with an AI in a public space. This could be due to a number of reasons. One participant who approached Rosie mentioned that it was very intimidating to start the conversation as the facilitators were sitting right behind Rosie and it felt like they would be "disturbed".

A modification of this experiment would be to remove the facilitators from the nearby area. Either be setting up a remote connection or by using room dividers to create privacy. As the room dividers allows the facilitators to still hear the users without delay this will be implemented for the second setup.

Rosie spends a lot of time talking in the beginning of the conversation while the really interesting part comes when she asks about current problems. This was more of a two-way conversation, opposed to the mundane yes/no answers from the structured part in the beginning. For this reason a part of her script is removed, which means that she no longer asks what the most interesting part of the project has been so far.

Additionally Rosie sometimes mishears participants as saying no to the consent form while they actually said yes or something that semantically resembles yes but isn't actually yes (e.g. "Go

ahead"). For this reason she no longer accepts a "no" and the facilitators will have to end the experiment manually before any user audio is recorded. These modifications are documented in appendix E.

The changes from setup 1 to setup 2 are:

- Room dividers to create privacy.
- Rosie's script is shortened and she no longer asks for the most interesting part of the project.
- The facilitators ends the experiment if the participant says no to consent.

5.3.5 Setup 2

The second setup implemented the changes from the previous discussion and the location was moved to the place that is marked as "Experiment 2" in figure 5.1. A picture of this setup can be seen on figure 5.5.

Figure 5.5: This figure shows a picture of the second setup. Rosie has been moved to a more secluded area that isn't in any major walking path. She is observable from multiple locations. The facilitators are located behind the room dividers

5.3.5.1 Results - Setup 2

Setup 2 was used for one session and had significantly fewer interactions. However, all interactions were successful. In total, four people interacted with Rosie. Many others showed interest by pointing her out to friends or glancing in her direction multiple times, but most chose not to engage.

Of the four participants, one mistook the experiment for another study they had already agreed to join, and another participated because they were actively looking for study participants themselves.

Due to storage issues, audio was only saved for the final participant. Additionally, the Wi-Fi failed during two interactions, meaning no data was collected about the participants' experiences in those cases. However, their initiation of the interaction was still recorded.

5.3.5.2 Discussion - Setup 2

Only a few people approached Rosie for the second experiment. This was much unlike the first setup where a total of ≈60 people were present in the area. This was likely a combination of fewer people present and that the experiment had moved physical location to be more secluded place. This led to people having to actively, make the decision, to approach Rosie. As such the second experiment went too far in the direction of not disturbing which meant that almost no participants were considered as participants. A third setup should try to find some middle ground between setup 1 and 2.

Additionally it is believed that many of the students at electronic systems AAU have seen Rosie but have not built up the courage to approach her. Therefore it is preferential to keep experimenting in the same general area as perhaps they have the courage to talk to her after 1-2 days.
One participant mentioned that they had seen Rosie both yesterday and earlier that day but didn't feel ready to interact. This supports our suspicion that more people may engage over the coming days as they build up the courage.

It was noticed by the facilitators that some participants would talk immediately after Rosie had stopped speaking, before the CLI tool writes "LYTTER" (Listening in danish), and as such the first part of their sentence would not be recorded. A solution to this could be to implement some asynchronous handling of the audio data or to not stop the recording until the user had repeated

themselves. To spare development time the second option will be tried in setup 3. This means that if users speak too fast such that the first part of their sentence isn't recorded, the facilitators will not press the "Stop recording" button on the GUI. This will result in the CLI interface still writing "LYTTER", hopefully prompting participants to repeat themselves.

The changes from setup 1 to setup 2 are:

- Use a more public setting.
- Wait to start the recording of participants until Rosie is ready.

5.3.6 Setup 3

Setup 3 was used for two sessions and took place in the location marked as "Experiment 3" in figure 5.1. The setup is also displayed in figure 5.6.

Figure 5.6: This figure shows a picture of the experimental area used for setup 3

5.3.6.1 Results - Setup 3

Setup 3 was used for two sessions. The first session had 9 of 21 successful interactions and the second session had 1 of 8 successful interactions.

It was observed that almost all conversations end after the first problem has been discussed and Rosie asks if there are any other things they would like to talk about.

Several times in this session, the program ran into an error where the returned audio from OpenAI was too big for the gRPC protocol leading to the program shutting down and therefore the interaction ended. These interactions were still marked as successful as the initiation happened correctly.

5.3.6.2 Discussion - Setup 3

In the user-initiated condition, several participants appeared unsure of how to begin interacting with Rosie and so they asked the facilitators. This behavior had not been anticipated during the experiment design phase. However, it posed a significant issue, as the two hypotheses required conflicting experimental strategies. Optimizing for H_1 necessitates minimal facilitator involvement to simulate a realistic scenario in which no personal guidance is available. In this case, users must either find their own way to initiate interaction with Rosie or choose to disengage entirely. In contrast, optimizing for H_2 requires that users successfully initiate a conversation to have more qualitative data, which could be encouraged by providing a simple instruction such as “Just say hello.”

In the moment, it felt inappropriate, to ignore participants. Consequently, the decision was made to assist them, thereby prioritizing H_2. However, this introduced a third outcome category: "Success – asked for help." This complicates the evaluation of H_1, as facilitator involvement alters what was intended to be a fully user-driven condition.

Participants who asked for help were noted in the observer logs, and this categorization is visualized in figure 5.7.

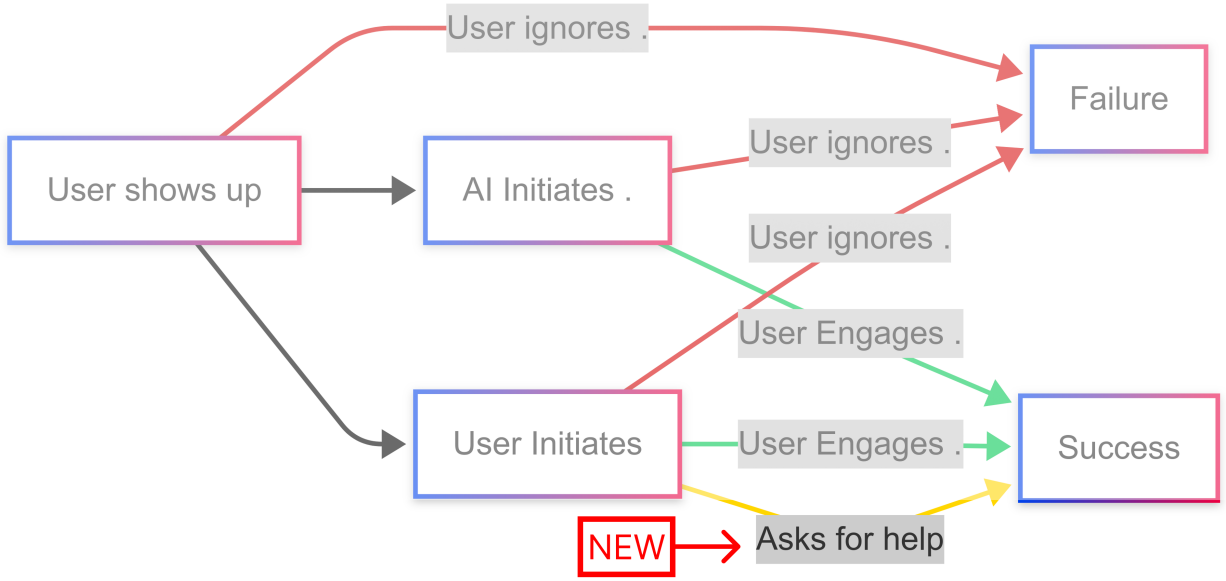

Figure 5.7: This figure shows a more detailed version of figure 5.2 with the addition of a new pathway to success in the user initiated condition called "Asks for help"

There are two erroneous conclusions made in the experiment. These are marked in purple in figure 5.8.

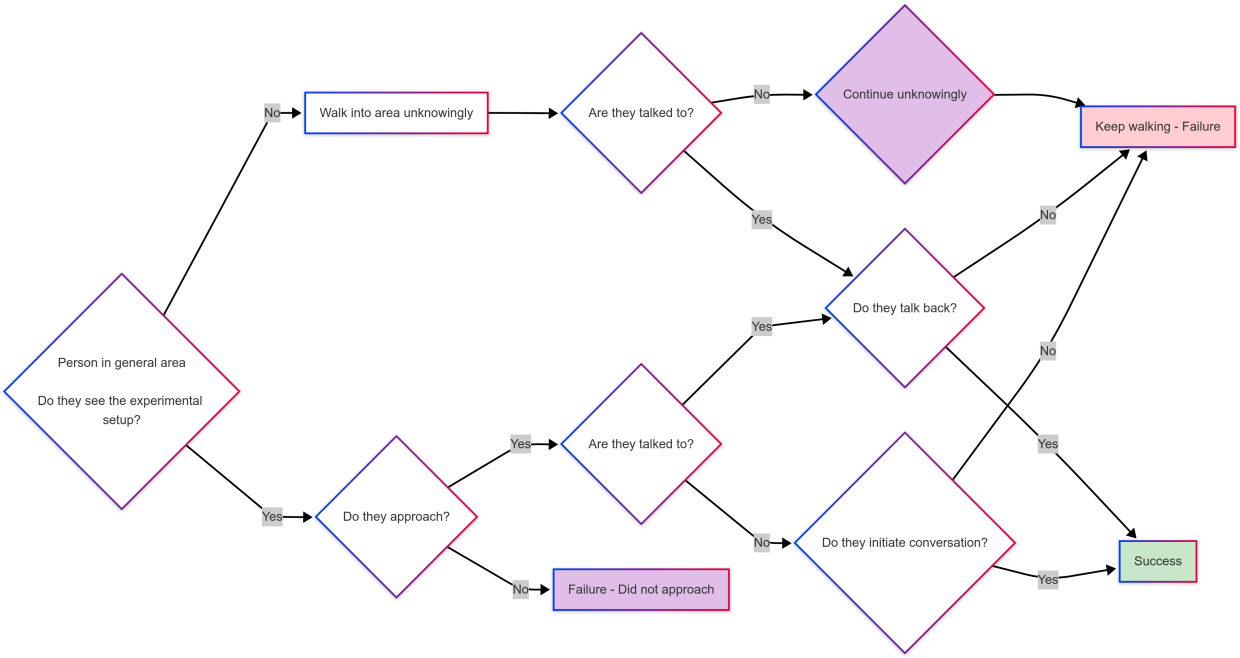

Figure 5.8: This flowchart shows a more accurate version of figure 5.2 with the separation of users either discovering that Rosie exists or not seeing her and as such being unable to make a choice to engage. It also has the addition of people seeing Rosie but not approaching and being to far away to have Rosie initiate the conversation. The two blind spots of the method are marked in purple

The first issue concerns users who unknowingly walk into the experimental area without noticing the setup. In such cases, the AI-initiated condition is disproportionately favored: if the AI initiates the conversation, users become aware of the setup and may choose to engage. However, if the AI does not initiate, these users remain unaware and thus are unable to make a choice. Until now, their behavior has been recorded as a failure in the user-initiated condition, which is misleading.
As such, in the fourth setup, users will only be included in the study if they clearly perceive Rosie, i.e., if they look at her at some point. Individuals who pass with their backs turned or are otherwise visually or cognitively disengaged will not be contacted or included in the dataset. However, people with headphones who look at her are included. While the audio prompt may be less effective for them, it is plausible that such users could remove their headphones in a user-initiated condition to engage.

Secondly, it was observed that some people observed Rosie from a greater distance - far enough for Rosie not to initiate a conversation. It could be argued that those people should be recorded as failures as they had their attention on Rosie but chose not to act. However, there will be no recording of this as it would skew the data and require the facilitators to be observant of the whole room.

The questionnaire was also revised after setup 3. The phrasing of the engagement-related items was adjusted to better reflect the subjective nature of the interaction. Specifically, the instruction was changed from “Please indicate how much you agree or disagree...” to “Please indicate to what extent the statement reflects your experience...”, aligning more accurately with the intention behind the engagement scale.

A showcase of the new personality as well as the experimental setup can be seen on YouTube[1] (Jeppe Givskud, 2025b).

The changes from setup 3 to setup 4 are:

- The facilitators will be better hidden and look "unreachable" to the participants.
- Bypassers will only count as part of the experiment if they are aware of Rosie.

[1] https://www.youtube.com/watch?v=uAh9NrMyKeE

5.3.7 Setup 4

As the students at Electronic systems seems to have all seen Rosie at this point it has been chosen to set up the experiment somewhere else.

For this purpose the fourth setup was done at Aalborg University Library Kroghstræde in a very open space with lots of traffic close to the entrance, the western section and the cantina. The setup can be seen in figure 5.9. Only one of the facilitator will sit near Rosie (behind the plant), the other two will be placed at a distance.

Figure 5.9: This figure shows the fourth setup. The facilitator is hidden behind the plant and the other researchers are placed away from the experiment

5.3.7.1 Results - Setup 4

Setup 4 was used for one session and has a total of 91 participants. All participants were aware of Rosie's existence and in close proximity to her. There were 10 successful interactions of which seven were AI initiated and one had an error with the Wi-Fi.

5.3.7.2 Discussion - Setup 4

The ratio of AI-initiated to user-initiated interactions in Setup 4 differed significantly from the previous setups. This disparity may be attributed to the highly public location of Rosie during this experiment. Initiating a conversation in such a setting could be perceived as intimidating, as it might draw attention or disturb people passing by. Furthermore, participants had to turn their backs to the main area of the room to engage with Rosie, which may have added to their discomfort. In this context, AI-initiated interactions may have played a crucial role in lowering the barrier to engagement, encouraging curious participants to respond despite the intimidating environment. This interpretation is further supported by the nature of the interactions since many occurred in groups, as illustrated in figure 5.10.

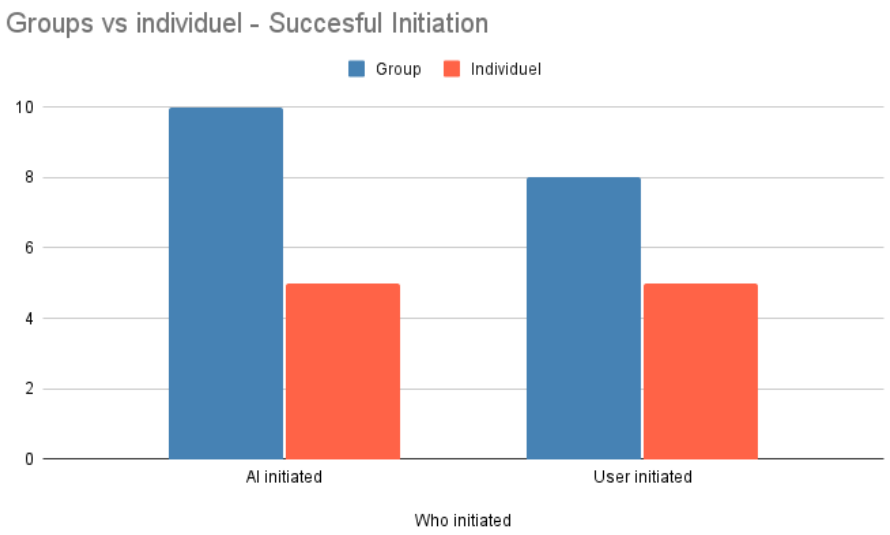

Figure 5.10: This figure shows the proportion of how many interacted with Rosie while in a group vs when alone for all four sessions. The data is from the observer notes

The facilitators were better hidden in setup 4 and only one participant asked for help. This was because the participant did not understand Danish.

Due to time constraints it was decided to not do more experiments and instead use the data from all four setups collectively.

5.3.8 Discussion of Setups

If users quickly moved across the area the program could not keep up. The program takes about 5 seconds to restart when a new condition is chosen (whenever a new user walks into the area). As such some users were missed.

The program and text surrounding the setup is all in danish. As such international students and staff that walk by may falsely contribute to a slightly higher failure rate.

Another observation from the experiment is that the seems to influence how willing people are to engage in the experiment. More participants seem to interact when the experiment is conducted later in the day.

5.4 Quantitative Results

The experiment was conducted in six separate sessions with four different setups across the university grounds in public locations. Session 1 and 2 was done with setup 1. Session 3 was done with setup 2. Session 4 and 5 was done with setup 3 and session 6 was done with setup 4.

The data has been gather and analyzed with focus on answering the two hypotheses H_1 regarding probability to participate in a conversation and H_2 regarding the subjective experience of the conversation.

The analyzes is made using Rstudio. The code can be found in attachments "Rstudio - Code" and on GitHub[2]

The data can be found in attachment "Quantitative data from experiment".

The analysis uses the following packages:

- tidyverse
- dplyr
- tidyr
- ggplot2
- Hmisc

[2] https://github.com/JeppeGivskud/OpenAIDigitalHuman/tree/master/RSTUDIO

5.4.1 Probability of success

A summary of all the participants outcomes over time can be seen on table 5.1 and figure 5.11.

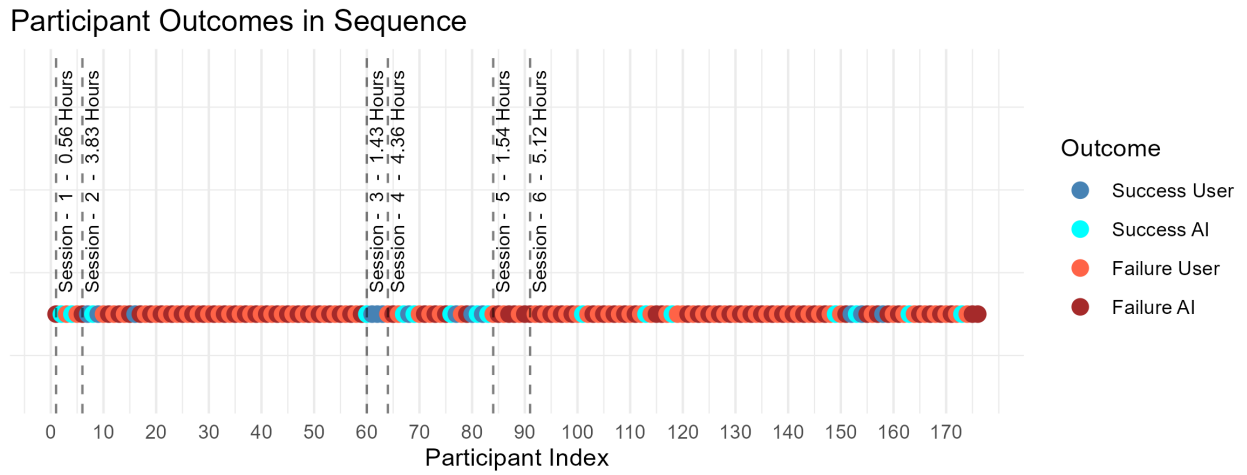

Figure 5.11: This figure shows all the six sessions and each participant and their results. The unknown results are when users leaves the experiment before finishing.

These summaries highlight how few people interacted with Rosie during each sessions as well as how different the sessions where in regards to total amount of participants.

Session lengths ranged from 0.56 hours to 5.12 hours, and the number of individuals passing through the experimental area varied considerably, from as few as 4 (Setup 3, Session 2) to 86 (Setup 4, Session 6).

Session	Setup	Session size	Session hours	Total Successes	Total Failures
1	1	5	0.56	2	3
2	1	54	3.83	4	50
3	2	4	1.43	4	0
4	3	20	4.36	9	11
5	3	7	1.54	0	7
6	4	86	5.12	10	76

Table 5.1: This table shows the results for each of the six sessions

The experiment had 176 participants in total. In the condition where the user initiated the conversation, there was 74 failures, and 14 successes in total. Out of those 13 successes, 7 participants asked for help. In the condition where Rosie initiated the conversation, there was 73 failures and 16 successes. Only 1 participant asked for help in this condition. This shows that there were many more failed than success and that the ratio of success between conditions are about the same. However the figure also shows how many participants asked for help in each condition and here the user initiated conditions had a lot more people asking for help.

The combined results can be seen below on table 5.2 and figure 5.12.

	AI Initiated	User Initiated
Failures	73	74
Successes	15	7
Success (Asked for help)	1	6

Table 5.2: This table shows the general results from the study

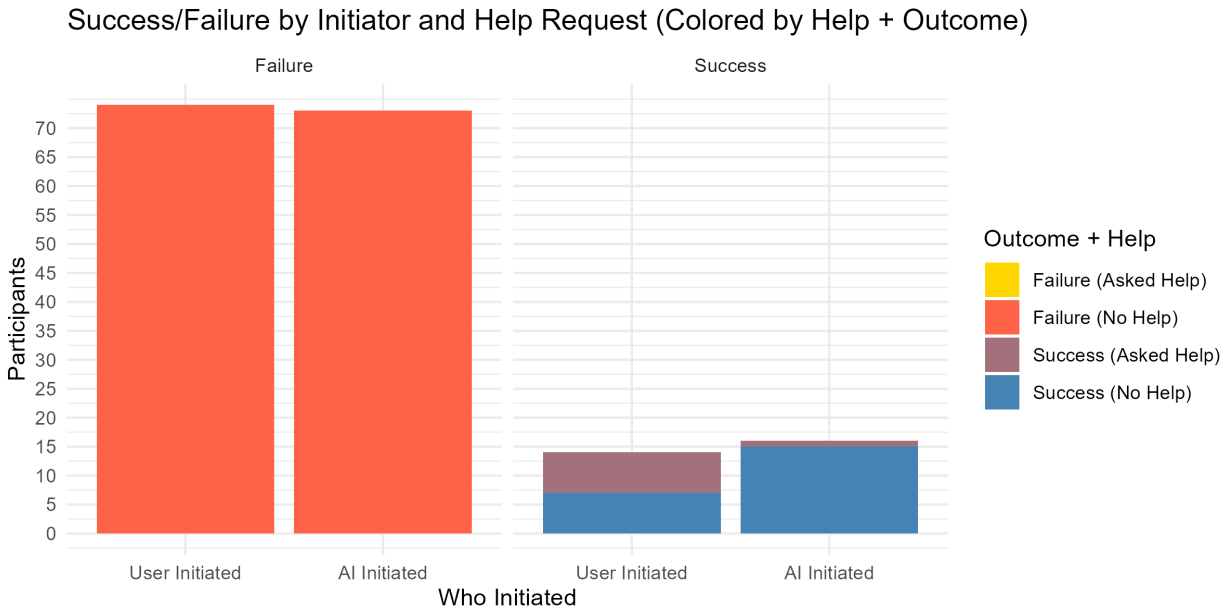

Figure 5.12: This figure shows the total count for each condition split into success and failure. Each column is also split into whether the user asked the facilitators for help on how to interact with Rosie.

To ensure comparability across experimental setups, all quantitative metrics were normalized due to significant variation in both setup duration and participant exposure.

These ratios are calculated as the proportion of successful interactions relative to the total number of observed individuals within a given setup. This approach ensures that variations in setup visibility, foot traffic, and time of day do not skew the interpretation of user engagement levels.

The normalized results can be seen on figure 5.13. The setups have a large variation in both amount of participants and the amount of AI to User success. The first and last setups have a lot of people per hour ca. 15 compared with the second and third setups which have ca. 4. This is likely due to these setups being located in more busy walking paths. Interestingly this doesn't seem to impact the amount of Successes per setup which is consistently low (0.7-2.1 per hour) and this value doesn't increase with the amount of failures. The second setup seemed to favor the user initiated condition, this is likely due to a methodological error as only 1 of 4 participants were presented to the AI initiated condition. As all users who walked into the area in setup 2 had successful interactions the outcome would likely have been equal if there had been no methodological error.

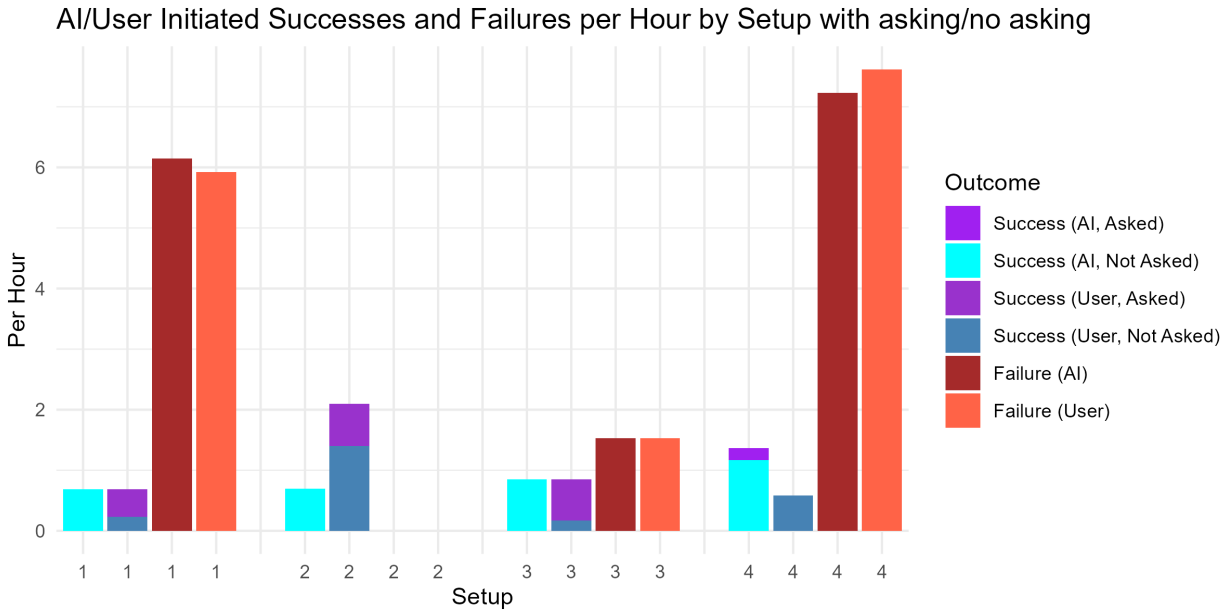

Figure 5.13: This figure shows the data with participants per hour. The x-axis shows each of the four setups while the y-axis shows how many participants fall into one of the four groups pr hour. The groups are failure vs success and AI vs User. Additionally the success group is split into "Asked for help" and "Did not ask for help".

The figure also illustrates how many participants asked for help during each session. Interestingly, the first three sessions show a high number of help requests for the user initiated condition, whereas the fourth session had no such requests. This is likely due to the facilitators being hidden during the fourth setup. Notably, this same setup also shows a higher number of AI-initiated successes. This

suggests that participants in setups 1–3 who asked for help might have otherwise given up had the facilitators not been visible or available, implying that facilitator presence heavily influenced the outcome.

To answer the hypothesis of whether there is a higher probability of speaking with Rosie when she initiates the conversation, a statistical analysis will be performed. Since the amount of participants in each condition that succeed is relatively low a Fishers exact test will be performed rather than the chi squared test as this test is better at handling few data points (Agresti, 2018, p. 212).

The test will answer whether there is a higher probability of speaking with Rosie if she speaks first.

The probability of Success as well as the Fishers exact test result can be seen below on table 5.3 and figure 5.14. The resulting p-value of 0.842 with odds ratio 1.246 indicate no statistically significant difference in probability to engage, depending on who initiates the conversation as such there is not enough evidence to reject $H0_1$.

	AI Initiated	User Initiated
Successes	16	14
Failures	73	74
Success Rate	0.180	0.159
p-Value	1	0.842
odds-ratio	1	1.246

Table 5.3: This table shows how often users participate in conversation with Rosie based on the two conditions AI initiated or User initiated. The p-value is calculated via Fisher's exact test.

Given the binary outcome (Success or Failure), it is also relevant to consider the absolute probability of success in each condition. Figure 5.14 shows that both conditions have a probability of success well below 0.5, indicating that participants were generally unlikely to engage with Rosie, regardless of who initiated the interaction. This suggests a general tendency toward non-engagement across conditions, rather than a difference in probability between them.

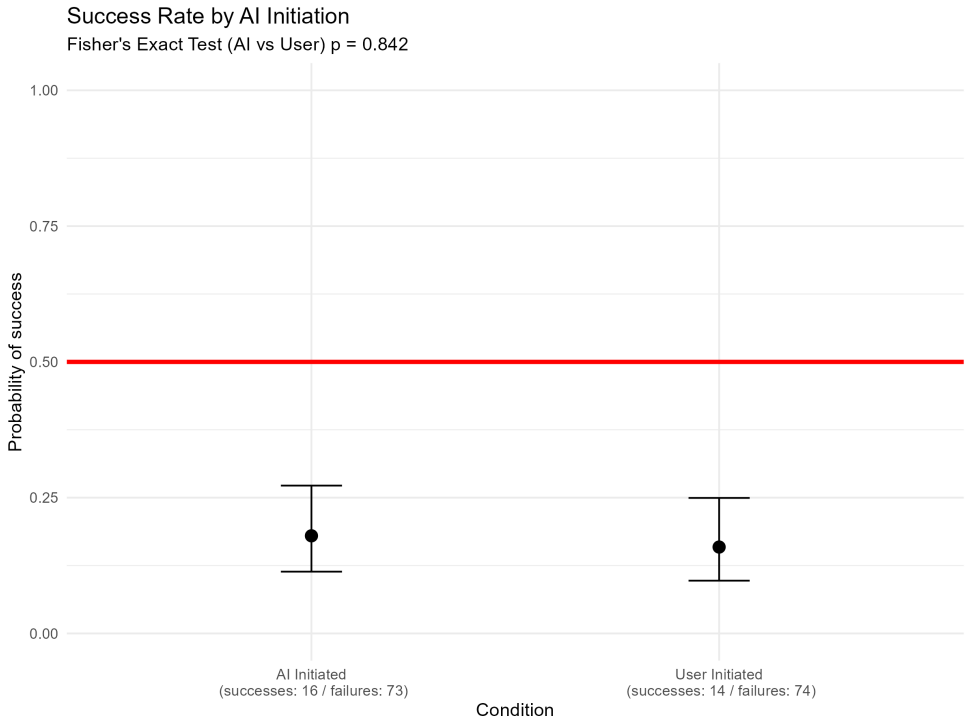

Figure 5.14: This figure shows how often users participate in conversation with Rosie based on the two conditions AI initiated or User initiated. Whiskers show confidence interval calculated as Wilson score interval.

However it is interesting to look into what the results could have been, had the facilitators not interfered with the experiment i.e. not been present for asking help. This is a pretty common outcome as shown on figure 5.13.

It is important to highlight that this analysis is very speculative and the results should at most lead to new studies examining the effect correctly.

To conduct this analysis, three interpretations have been made of the groups who asked for help:

- The users who asked for help would succeed no matter what and should stay in the succeed group.
- The users who asked for help should be marked as errors and removed from the data.
- The users who asked for helped would not have been able to succeed without it and should be marked as failures.

To test how these interpretations affects the probability of participating in a conversation, each of the groups have been compared to the group where Rosie initiated the conversation with a Fisher exact test. Table 5.4 and figure 5.15 shows that across all interpretations of the failures, the probability

of success remains under 0.5. This suggests a general tendency toward non-engagement no matter condition or interpretation.

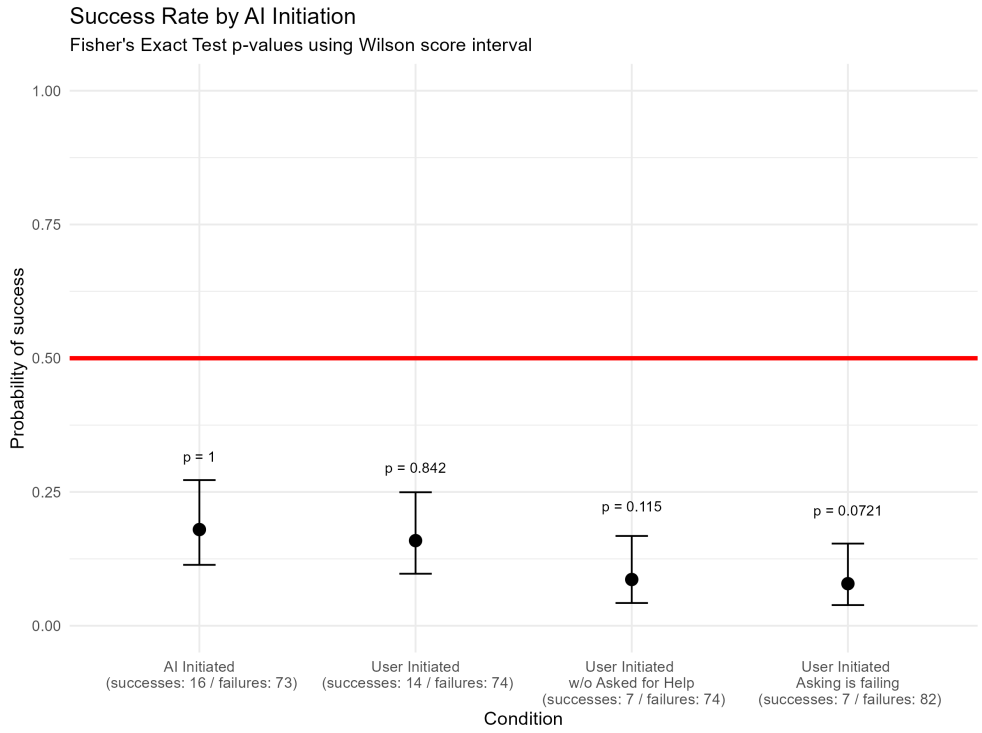

Figure 5.15: This figure shows the success rate for the AI initiated condition as well as for the three assumptions leading to three different user initiated conditions. The first is the assumption that all users who asked for help succeeded. The second is the assumption that users who asked for help should be excluded. The third is that all users who asked for help would have failed. Whiskers show confidence interval calculated as Wilson score interval.

However a trend towards a lower p-value is observed. As such it could be speculated that if more data was gathered where users are not able to ask for help the result would then show. This is further supported by the fourth setup having fewer ´User successes´ than ´AI successes where nobody asked for help´ as the facilitators were hidden.

	AI Initiated	User Initiated	User Initiated w/o Asked for Help	User Initiated Asking is failing
Successes	16	14	7	7
Failures	73	74	74	82
Success Rate	0.180	0.159	0.086	0.079
p-Value	1	0.842	0.115	0.072
odds-ratio	1	1.246	2.306	2.523

Table 5.4: Performance statistics across different conditions or assumptions: AI initiated, User Initiated, User Initiated w/o Asked for Help, User Initiated where asking is failing. The P-values are from fishers exact test.

5.4.2 Probability of asking for help

Another result is that the data indicates that users are more likely to ask the facilitators for help if they are to initiate the conversation. To test whether this is the case or not, it needs to be treated as a hypothesis. The study then investigates three hypothesis: *"Probability to engage in conversation"*, *"Probability for a different subjective experience of the conversation"* and *"Probability to ask for help"*. Therefore, it is required to do a correction for multiple comparisons to control the family-wise error rate. For this purpose, the Bonferroni correction has been applied. This adjusts the significance threshold to:

$$\alpha = \frac{p}{hypothesis} = \frac{0.05}{3} = 0.0167$$

.

As shown in Table 5.5 and figure 5.16, the resulting p-value of 0.012 falls below the adjusted threshold of 0.0167, indicating that participants were significantly more likely to ask for help when in the user initiated condition.

	AI Initiated	User Initiated
Asked	1	7
Didn't ask	15	7
Asked Rate	0.063	0.500
p-Value	1	0.012
odds-ratio	1	0.085

Table 5.5: This table shows how often users asked for help in the two conditions AI initiated and User initiated. The p-value is calculated via Fisher's exact test.

This finding is further supported by the fact that the user-initiated group's success rate is exactly 0.5, while the AI-initiated group's rate remains well below 0.5. This suggests a very low likelihood of asking for help when Rosie initiates the conversation, and no clear tendency — neither toward asking nor not asking — when participants initiate the conversation themselves.

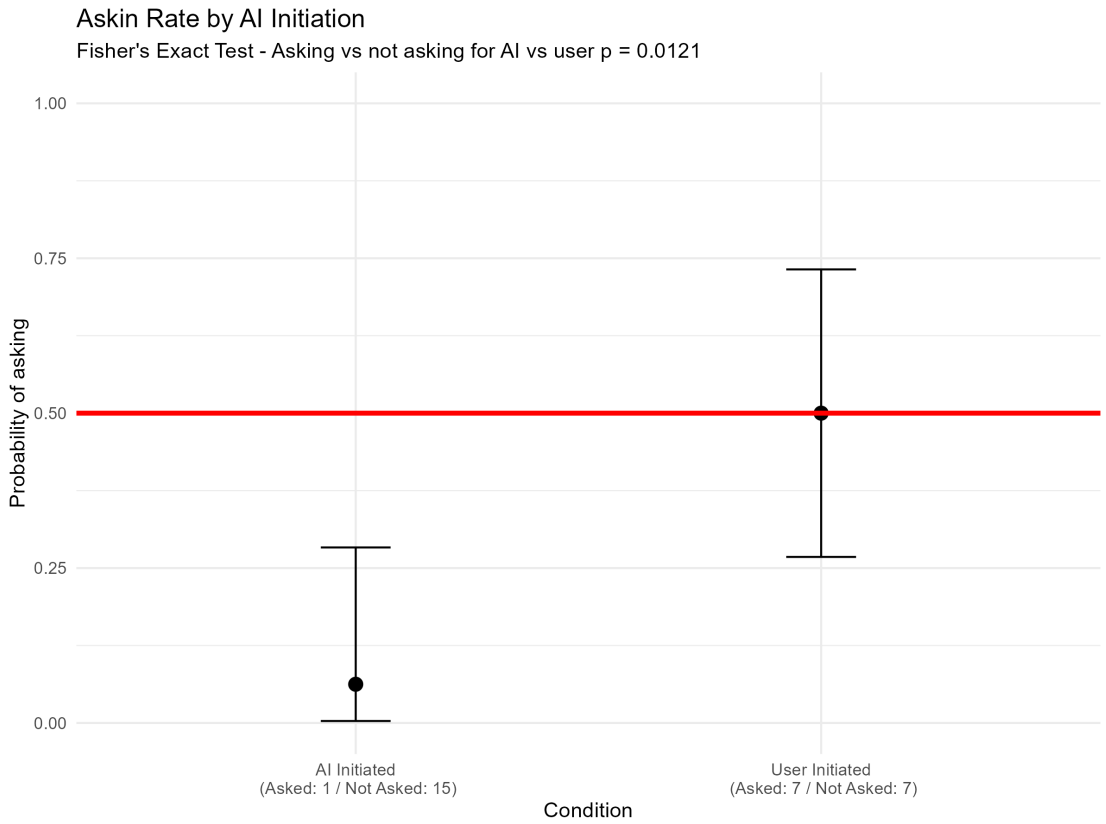

Figure 5.16: This figure shows how often users asked facilitators in the two conditions AI initiated or User initiated. Whiskers show confidence interval calculated as Wilson score interval.

5.4.3 Analysis of sentence data

One of the proposed ways in section 5.3.3, to measure a difference of conversation quality, was to investigate the changes in sentence length, total amount of turns and total conversation length. The audio was transcribed using Whisper AI and analysed using python scripts and Rstudio. The analysis can be found in attachments "Rstudio - Code - Transcription" and on GitHub[3].

Due to various technical issues not all of the interactions were recorded and transcribed. As such the following data is based on 22 out of the 30 conversations.

Boxplots of the three analysis can be seen below on figures 5.17, 5.18 and 5.19.

Figure 5.17 shows a large standard deviation in sentence length between participants $\approx$ 34 characters. This means that some participants have very short answers whereas others have very long answers.

[3] https://github.com/JeppeGivskud/OpenAIDigitalHuman/tree/master/RSTUDIO

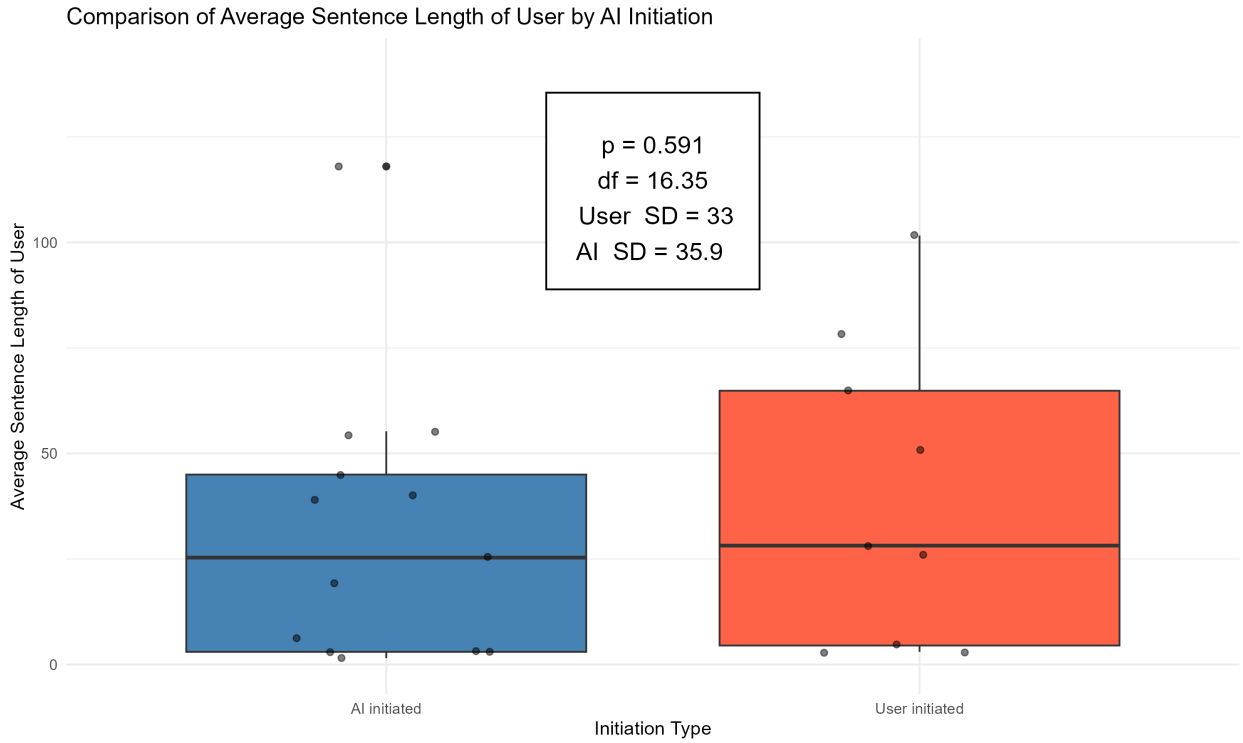

Figure 5.17: This figure shows a box plot of the average sentence length per participant grouped by condition AI initiated and User Initiated. Whiskers show inter quartile range. The dots are individual participants.

Figure 5.18 shows a large standard deviation in total conversation length between users $\approx$ 90 seconds. Some users have very short conversations < 1 minute while others speak for over 5 minutes. The users who speak less than 1 minute are perhaps due to setup errors where Rosie fails for various reasons.

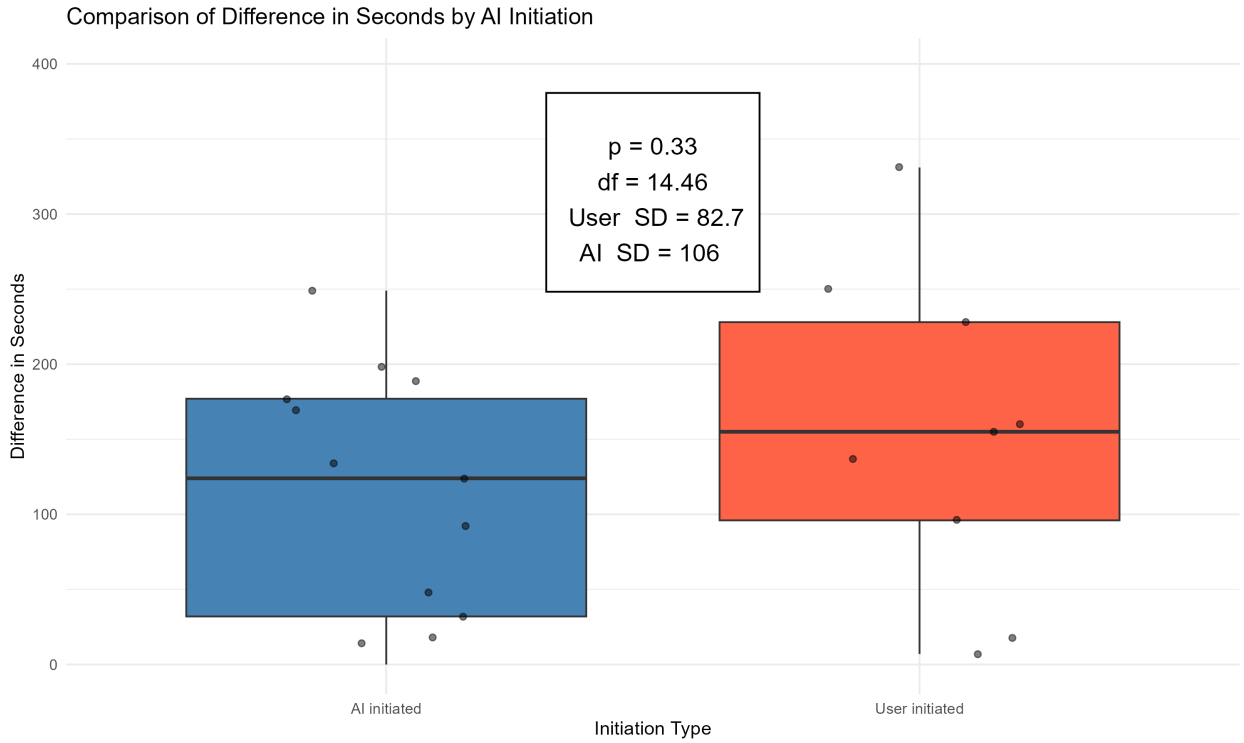

Figure 5.18: This figure shows a box plot of the average conversation length per participant grouped by condition AI initiated and User Initiated. Whiskers show inter quartile range. The dots are individual participants.

Figure 5.19 shows a smaller standard deviation in total turns between each participant $\approx$ 7.5 turns. That is, most conversations were around 5-15 turns. This is likely due to Rosie having a very strict structure of asking for consent which is 3-4 turns. Then asking about the project in 1-2 turns, and then any difficulties 1-5 turns.

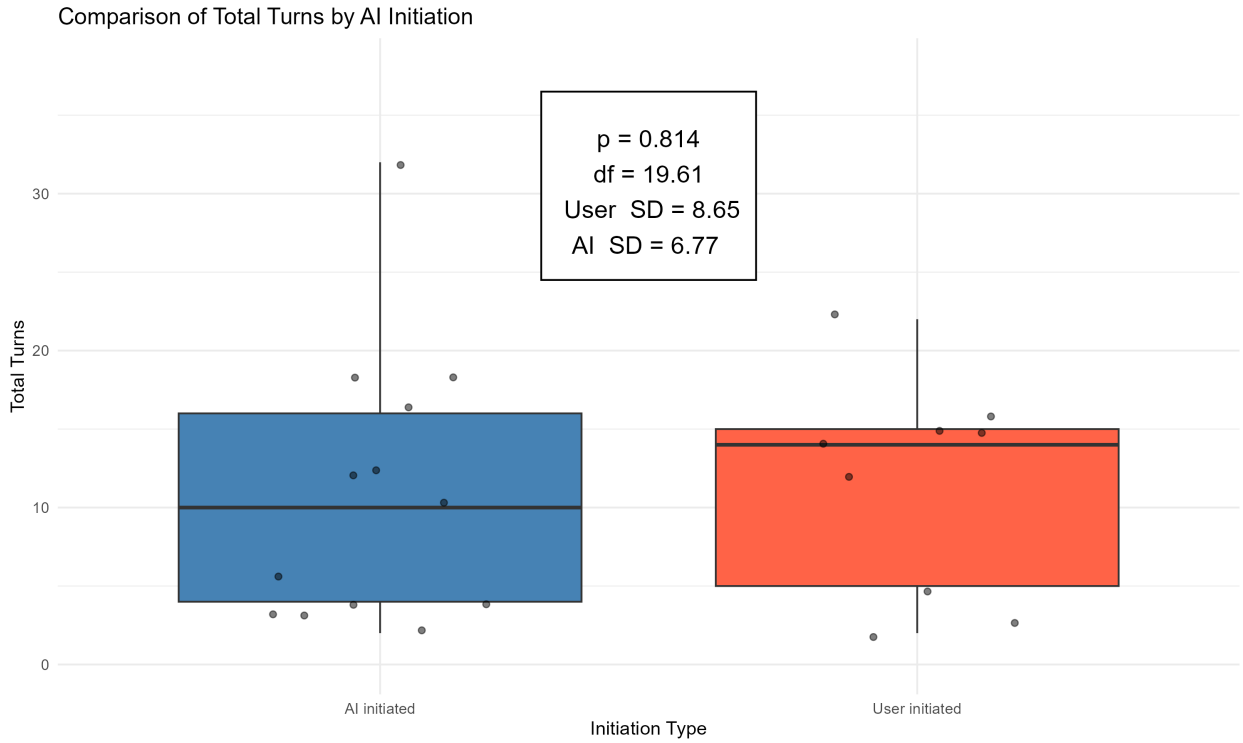

Figure 5.19: This figure shows a box plot of the total amount of turns per participant grouped by condition AI initiated and User Initiated. Whiskers show inter quartile range. The dots are individual participants.

For each variable: Sentence Length, Conversation length and total turns an independent two-sample t-test has been conducted and the results can be seen below on table 5.6. None of the resulting p-values are below 0.05 and as such there is no reason to reject the null hypothesis of significant differences in conversational behavior between user initiated and AI initiated conversations.

	t	df	p-Value
Sentence Length	-0.548	16.348	0.591
Conversation Length	-1.009	14.456	0.329
Total Turns	-0.239	19.611	0.814

Table 5.6: This table shows the results of the independent two sample t-tests for sentence length, conversation length and total turns between the two conditions AI initiated and User initiated

As it was observed during testing that users initially spoke in very short and simple sentences to Rosie, an analysis was conducted to examine how sentence length changed over the course of each conversation. The results are shown in figure 5.20. Sentence lengths tend to be very short in the first few turns, likely because users were responding to yes/no questions. Around the third or fourth turn,

participants are asked about their project, which leads to much greater variation in sentence length. In general, sentence length appears to increase after this question, suggesting that users elaborate more once she engages with their input. However, this pattern is not consistent across all participants. Some users speak at length early on and then become more concise. Most conversations end with short sentences, likely due to users declining further assistance with brief responses such as "no".

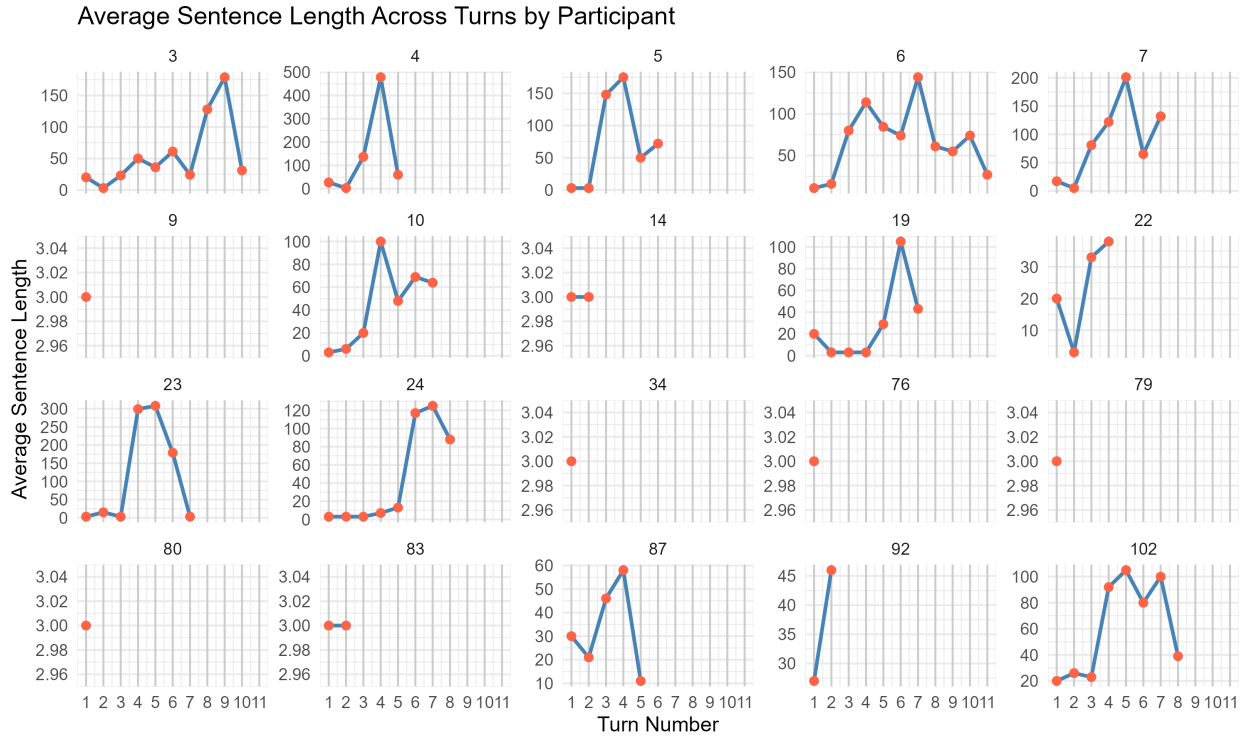

Figure 5.20: This figure shows the sentence length per turn for each of the 22 participants. The x-axis shows turns and the y-axis shows sentence length.

The average sentence length per turn can be seen on figure 5.21. This figure shows a similar tendency of short sentences the first 1-3 turns and then longer sentences until ending in shorter sentences.

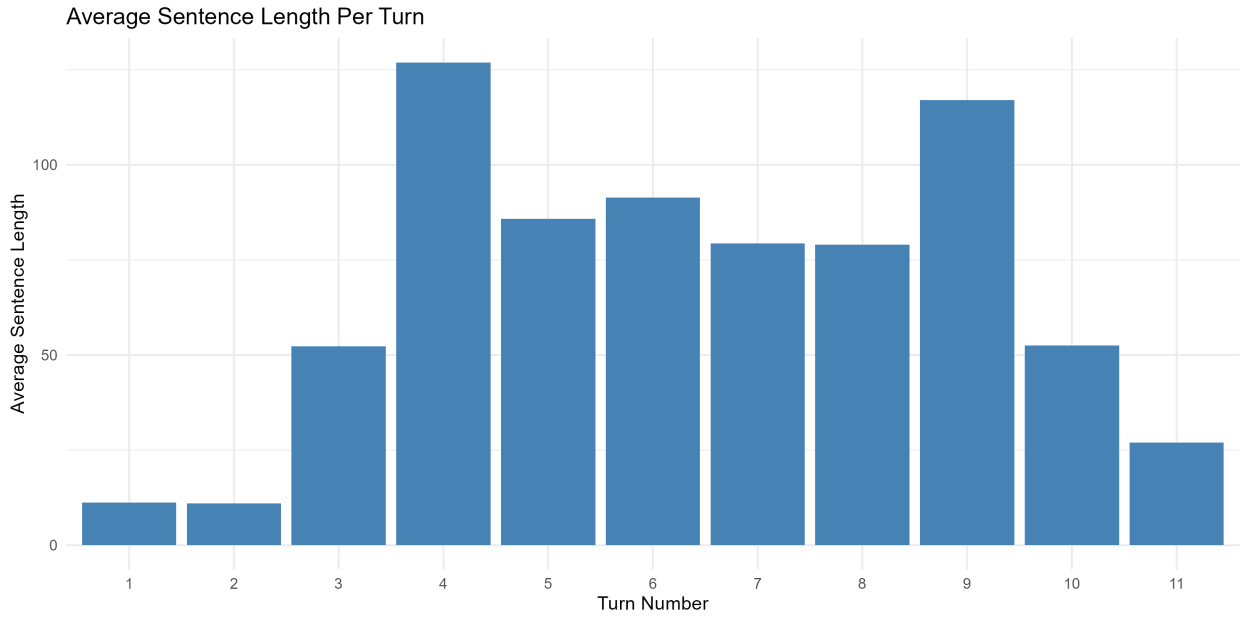

Figure 5.21: This figure shows the average sentence length per turn for the participants. The x-axis shows turns and the y-axis shows sentence length.

5.4.4 Analysis of Engagement and intention to use

Upon finishing the conversation with Rosie, the participants were asked to fill out a questionnaire containing subscales from TAM and TPU. These data are used to answer H_2. Participation was voluntary and participants could complete it on their own. In total 30 participants were offered the questionnaire and 18 participants completed it. The participants who completed were all in the age group 21-30 and were mostly males (13 males to 4 females to 1 unspecified).

All participants had interacted with generative AI's at some point as seen on figure 5.22. Most participant have used ChatGPT which is also the part of Rosie's that process the conversation.

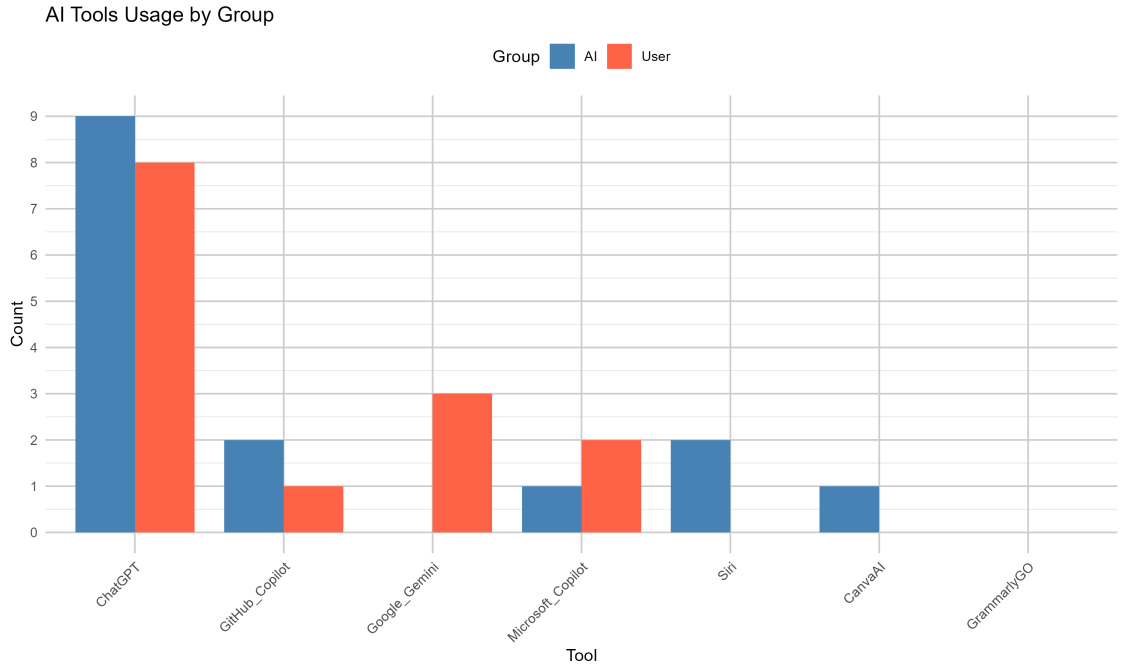

Figure 5.22: This figure shows which generative AI's the participants who completed the survey have interacted with grouped on condition (AI initiated vs User initiated).

To answer H_2 about changes in the quality of interaction if Rosie initiated the conversation three different constructs were measured: Perceived Ease of Use (PEU), Perceived Usefulness (PU) and Engagement. Each construct had six questions which were rated on a 7 point Likert scale and the average rating for the individual questions can be seen in appendix F.

The average rating for each of the constructs can be seen below on figure 5.23.

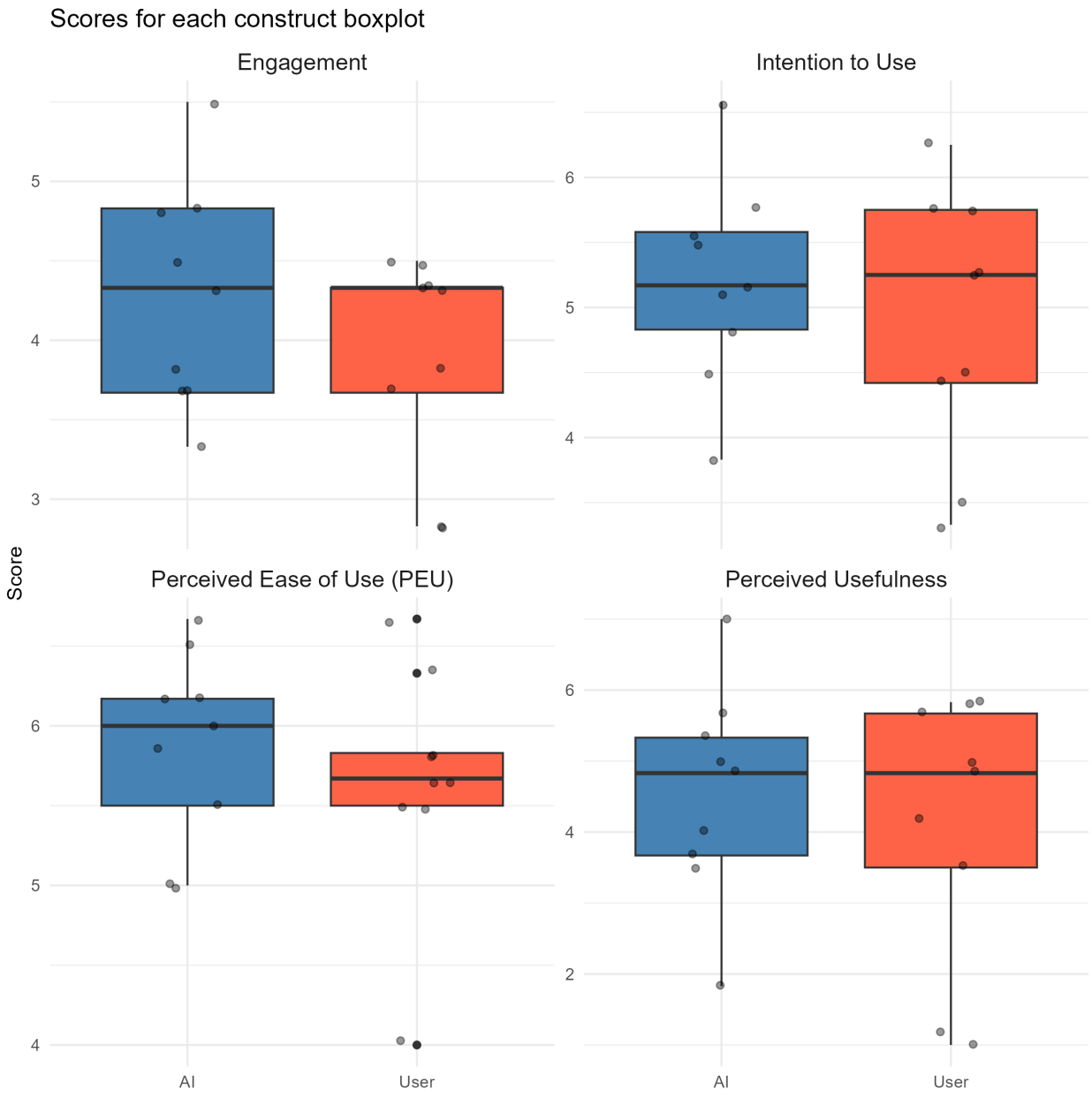

Figure 5.23: This figure shows the final score for each of the three sub-construct based on the 6 x 3 questions in the questionnaire as well as the final score for intention to use. The results of the intention to use and engagement have been tested via a two sample t-test and the results are not significant.

The average scores of perceived Ease of Use and Perceived usefulness is the score for Intention to Use (Davis, 1989; Park, 2009; Venkatesh & Davis, 2000) which comes out at 4.89 (AI initiated) and 5.2 (User initiated).

The average scores for AI initiated conversations are in general about 0.2-0.4 points higher on all the four constructs as seen on table 5.7 and figure 5.23.

	Engagement	Intention to use	Perceived Ease of Use	Perceived usefulness
AI Initiated	4.28	5.2	5.87	4.54
User Initiated	3.91	4.89	5.67	4.11
t	1.1	0.7	0.6	0.5
df	16	15	15.4	15.2
p-value	0.271	0.477	0.528	0.602

Table 5.7: This table shows the mean score for the four constructs measured in the questionnaire: Engagement, Intention to Use, Perceived Ease of Use and Perceived usefulness as well as the four independent two sample t-tests.

None of the results are statistically significant and as such there is no reason to reject $H0_2$.

5.5 Thematic Analysis

The following thematic analysis is based on observational notes collected during the experiment. These notes were taken in real-time and reflect participants' spontaneous comments and perceived reactions while interacting with the digital human interface. In order to organize and interpret the qualitative data meaningfully, the material has been coded and grouped into overarching themes that capture patterns in user experience and perception.

The thematic analysis is based on data from 20 participants, with 9 in the Rosie-initiated condition and 11 in the user-initiated condition. An additional 8 participants were excluded from the analysis. These were excluded for the following reasons: three spoke only English and could not engage meaningfully with Rosie, two reported not having time to interact, one greeted Rosie briefly and walked away, one stood in front of Rosie waiting for something to happen (but left when nothing occurred, as this was in the user-initiated condition), and one interaction was disrupted due to a network failure. Notably, 6 out of the 8 exclusions occurred in the Rosie-initiated condition.

To account for differences in interaction, the findings are divided into two distinct conditions: one in which the digital human (Rosie) initiates the conversation, and another in which the user takes the initiative to interact. This distinction enables a more nuanced understanding of how the direction of initiative influences participants' attitudes, comfort levels, and perceptions of the interaction.

The quotes in this section are translated from Danish to English by the authors and serve to back up analytical points. The translations have not been independently validated, but aim to retain the original tone and intention of the participants.

The thematic categories presented below serve to highlight key user perspectives and recurring affective or cognitive responses observed across both conditions.

The entire analysis and observation data can be found in attachment "Qualitative data from experiment".

5.5.1 Analytical Approach

To structure the analysis, we applied a reflexive thematic analysis following Braun and Clarke's six-phase framework (Braun & Clarke, 2006). The goal was to identify meaningful patterns across participant responses and support interpretation of the lived user experience with Rosie.

Familiarization: The research team read through all written responses and observational notes multiple times to gain a comprehensive understanding of the data.

Generating Initial Codes: Each participant's answer to the three open-ended interview questions was coded separately. Coding was conducted inductively, and most codes were constructed using direct quotes to remain close to participants' phrasing. Observations were treated as supporting context.

Searching for Themes: Codes were clustered into preliminary themes based on shared meaning. Responses were tracked across both user-initiated and AI-initiated conditions, allowing the researchers to trace emerging patterns specific to each interaction type.

Reviewing and Refining Themes: Themes were reviewed to ensure internal consistency and conceptual clarity. Subthemes were created where relevant to reflect distinctions within broader categories (e.g., Unrealistic Tone, Public Discomfort, Technological Curiosity).

Defining and Naming Themes: Themes were given descriptive names, with illustrative quotes translated from Danish to English, such as: "Fine" (original: "Fint") "She overpraises a bit" (original: "Hun overroser lidt") "It felt a bit weird to stand and talk to her in public" (original: "Det var lidt weird at stå offentligt og snakke med hende")

Producing the Analysis: The final themes are presented with linked participant quotes to support each interpretation. The analysis highlights how tone, initiative, and environment shaped the perceived authenticity, usefulness, and comfort of the AI interaction.

5.5.2 First Impressions by Interaction Type

The procedure for the thematic analysis was followed for each of the three exit questions, including the observations. First off is the participants' first impression of Rosie. The themes for both conditions can be seen in figure 5.24 and 5.25.

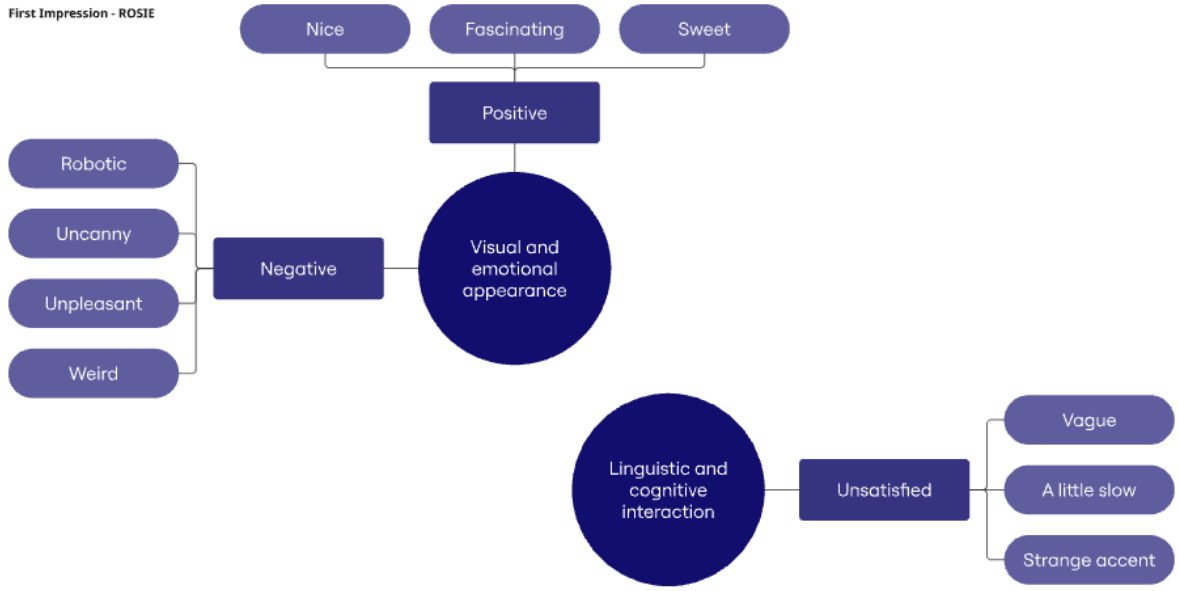

Figure 5.24: This figure shows the themes that describe the participants' first reaction to Rosie, when she initiates the conversation.

When Rosie initiates the conversation, the two major themes are *Visual and emotional appearance* and *Linguistic and cognitive interaction*, which is also displayed in figure 5.24. The subthemes for Rosie's appearance are *Positive* and *Negative*. Many noted her visual presentation as strikingly artificial. Descriptions such as "robot-like" and "uncanny" were common, and some participants found her both aesthetically impressive and fascinating.

The subtheme for the interaction is *Unsatisfied* as her responses were occasionally described as vague or overly slow. Her accent was also a point of critique, being perceived as mismatched or unnatural.

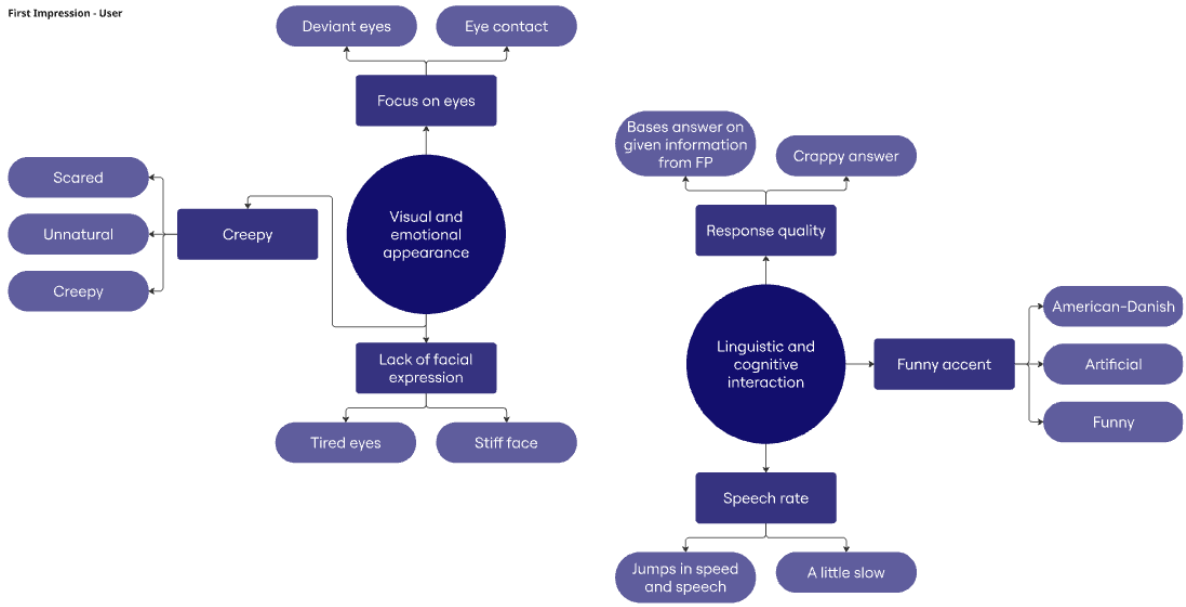

Figure 5.25: This figure shows the themes that describe the participants' first reaction to Rosie, when they initiate the conversation.

As seen in figure 5.25, the same main themes are present when the user is initiating the conversation but with different subthemes. When the user is initiating the conversation, the subthemes for *Visual and emotional appearance* are *Focus on eyes, Creepy*, and *Lack of facial expression*. Several participants focused on Rosie's lack of facial expressiveness, describing her as "stiff," "unnatural," or "a bit creepy." Some emphasized her eyes as "indifferent" or "tired-looking".

The subthemes for *Linguistic and cognitive interaction* are *Respons quality, Funny accent*, and *Speech rate*. While some were surprised at how well Rosie understood input, others found her speech pacing confusing or uneven. Descriptions included "she talks like Queen Mary — creepy," and "the accent was artificial".

5.5.2.1 Comparative Observations

Across both conditions, the same two main themes are central to the participants' first impression of Rosie. However, the emphasis within these themes differed depending on who initiated the interaction.

When users initiated the conversation, the participants were more likely to focus on Rosie's capabilities, performance, and quality. In contrast, when Rosie took the initiative, the participants focused more on superficial qualities.

In both conditions, participants paid close attention to visual realism and speech patterns. A recurring tension between curiosity and discomfort was evident, shaped not only by the design of the digital human but also by the physical and social environment in which the interaction occurred.

5.5.3 What made you stop and talk to her?

The participants was also asked why they stopped to talk to Rosie. The themes are divided into the two conditions and can be seen in figure 5.26 and 5.27.

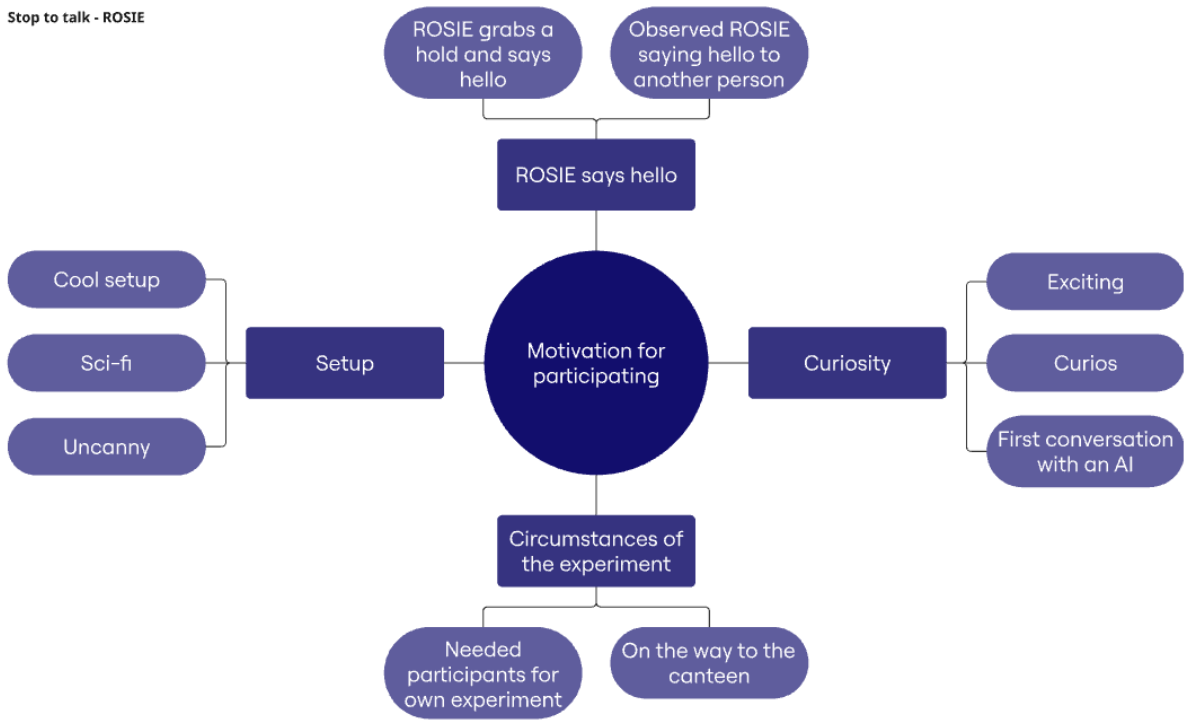

Figure 5.26: This figure shows the themes that describe why participants' stop to talk when Rosie initiates the conversation.

Figure 5.26 shows that when Rosie initiates the conversation, the main theme is that participants have a *Motivation for participating*. The subthemes and thereby their motivations for participating are *Rosie says hello, Curiosity, Circumstances of the experiment*, and the *Setup*.
Participants referred to Rosie's physical presence and her ability to initiate contact, such as "She stands here and grabs your attention by saying hi". Other participants were curios or simply intrigued by the setup and described it as immersive or futuristic. Some also mentioned the context like We

were missing participants for our own experiment, so we saw the sign," or "We were just going to see what was in the cafeteria".

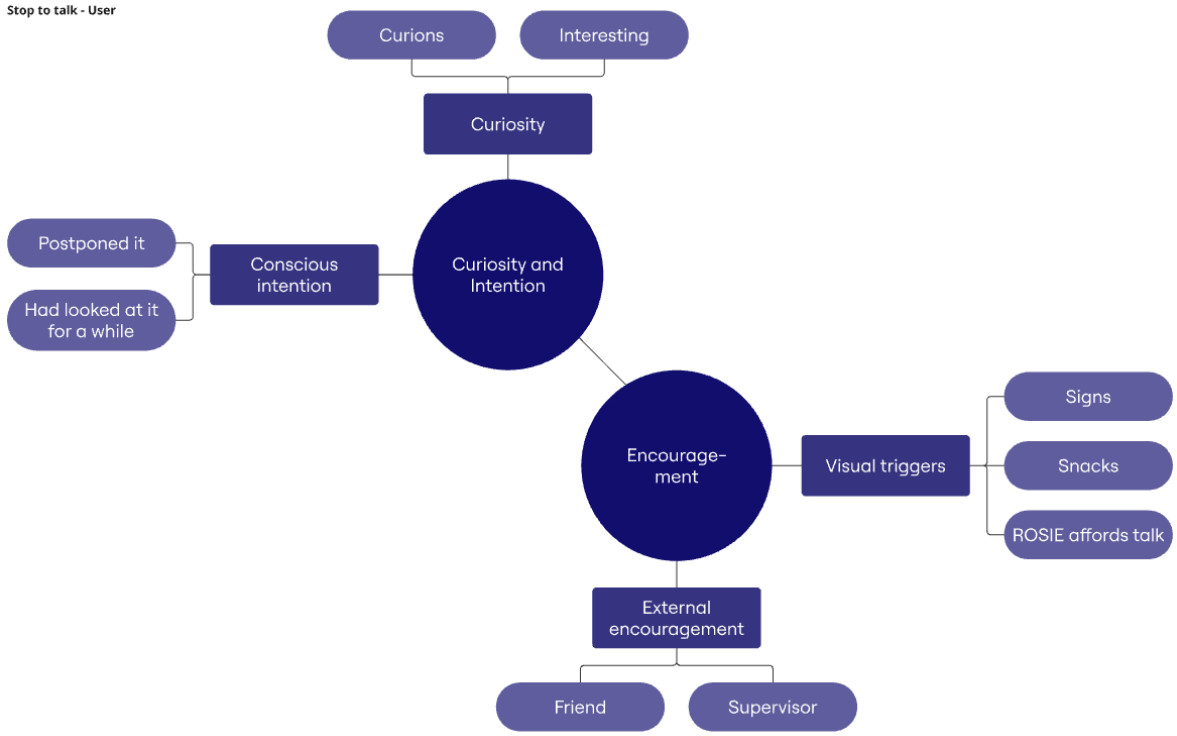

Figure 5.27: This figure shows the themes that describe why participants' stop to talk to Rosie, when they are initiating the conversation

When the user initiated the conversation, the main themes for why they stopped to talk is *Curiosity and intention* and *Encouragement*, which can also be seen in figure 5.27. Within the theme of *Curiosity and intention*, the subthemes are *Curiosity* and *Conscious intention*. Several participants described building interest over time "I had scoped it out already, I knew I wanted to try it" or were simply curious "I was curious because of the screen and the animation". or "I had kind of decided to do it later".

The subthemes for *Encouragement* are *External encouragement* and *Visual triggers*. These participants were influenced by social cues and subtle encouragement. For some, it was a nudge from a friend, "A friend encouraged me", or visual prompts and contextual cues, such as signage and Rosie's presence itself: "Because it was just standing there to talk to," and "There was a sign saying 'talk to me,' and I thought it could be fun".

5.5.3.1 Comparative Observations

Across both conditions, participants' motivation to engage was influenced by a mix of curiosity, contextual opportunity, and external triggers. However, the form of initiation shaped how these factors played out.

In Rosie-initiated interactions, engagement was often described as a reaction to being approached. The participants motivation was largely shaped by curiosity, immediacy, physical presence, and the concept of unsolicited contact, especially when they had not planned to interact.

In user-initiated interactions, engagement appeared more intentional, often supported by social reinforcement, design cues, or prior observation. Together, these responses highlight the importance of external triggers, both social and visual, in prompting users to initiate interaction with an unfamiliar AI in a public space.

In both cases, environmental design and social framing played key roles in shaping engagement whether by sparking attention or enabling the user to act when ready. These findings highlight that both passive exposure and active invitation can prompt interaction, but they activate different types of motivation and user agency.

5.5.4 How would you say the conversation went overall?

In the last question of the exit interview, the participants were asked to evaluate the conversation. The themes for these descriptions can be seen in figure 5.28 and 5.29.

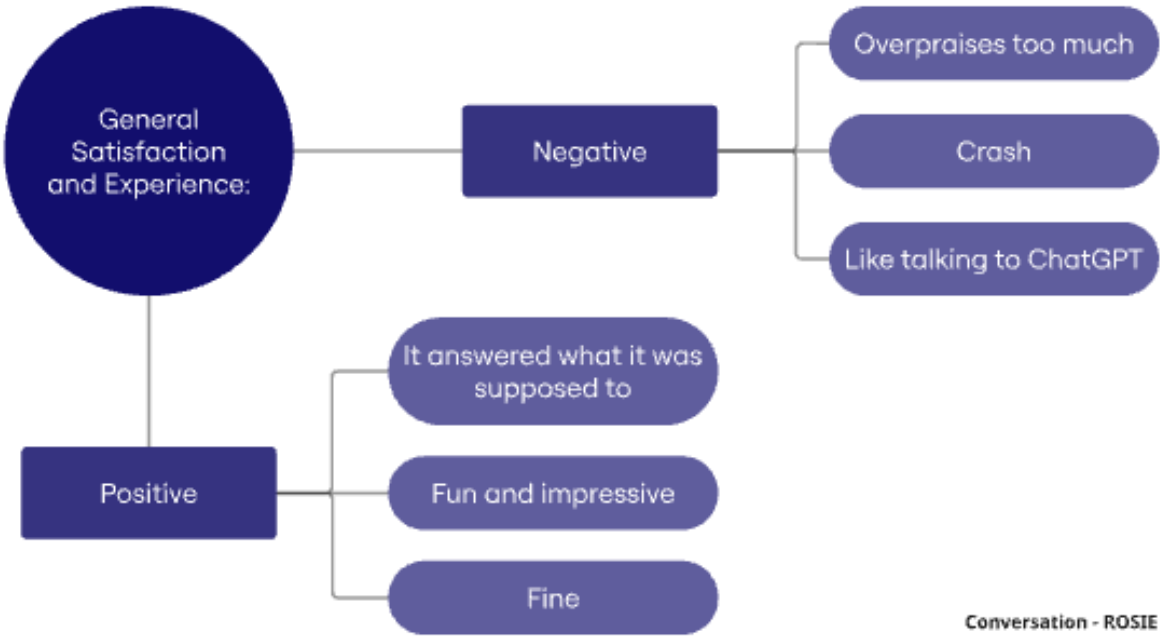

Figure 5.28: This figure shows the themes for how participants evaluate the conversation when Rosie initiates the conversation.

When Rosie initiates the conversation, the main theme for the evaluation is *General satisfaction and experience*, with the subthemes being *Positive* and *Negative*. As seen in figure 5.28 the positive evaluations include "Fine", “It went really well, it felt very natural”, or “It was fun, it was impressive”. Negative evaluations include “it went okay until it didn’t anymore” (crash) and "it has a bit of the classic American attitude, ’no, how cool’ - a bit too over-the-top happy. It overpraises a little".

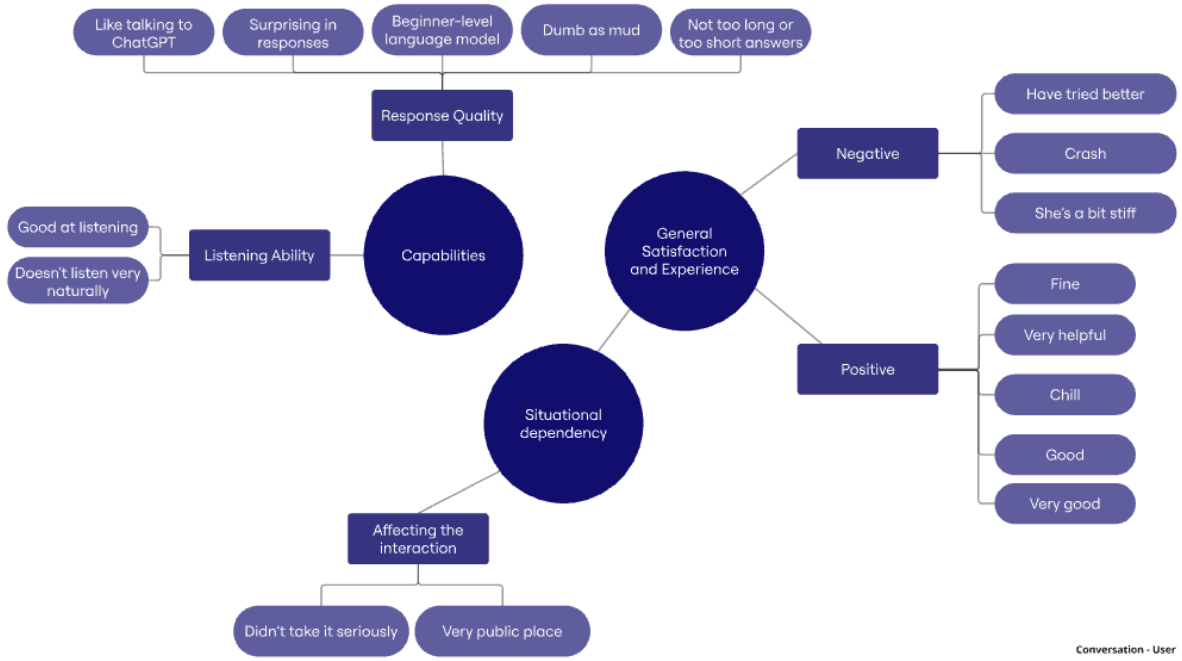

Figure 5.29: This figure shows the themes for how participants evaluate the conversation, when they are initiating the it.

For user-initiated conversations, the main themes for the evaluations are *Capabilities, Situational dependency*, and *General Satisfaction and experience*. This is also shown in figure 5.29.

For *Capabilities*, the subcategories are *Listening ability* and *Response quality*. This is described with comments such as “She was kind but dumb as a rock”, "It went pretty chill, she’s a good listener" and "It went well, she surprised me with her answers and how much she actually understood. She was very nice".

Affecting the interaction is the only subtheme within *Situational dependency*. This theme emerges from comments like “It’s a very public place to stand and talk to her” and "We didn’t take it that seriously”.

General Satisfaction and experience contains the subthemes *Positive* and *Negative*. The associated words to these themes are "Fine", "Helpful, "Good", "Have tried better", and "She’s a bit stiff".

5.5.4.1 Comparative Observations

Across both conditions, participants generally described the experience as positive, manageable, and occasionally surprising. However, subtle differences emerged in how the interaction was evaluated depending on who initiated it.

In Rosie-initiated interactions, users focused on flow and general satisfaction, with relatively few critical remarks. These participants were more likely to describe the experience in broad or affective terms of how it felt.

In contrast, user-initiated participants often gave more detailed evaluations, reflecting on intelligence, helpfulness, and conversation dynamics. This group seemed more invested in assessing Rosie's competence, and they more frequently pointed out technical limitations, unnatural behavior, or a lack of depth in responses. They also appeared more sensitive to the social environment, especially the public setting, as a factor influencing interaction comfort.

Together, the findings suggest that AI-led interactions may be judged more on tone and first impression, while user-led interactions invite more reflective evaluations possibly due to greater user intention and involvement in initiating the exchange.

5.5.5 Observations

The procedure for the thematic analysis was also followed for the observations that were noted during the experiment. These observations have also been divided into the condition where Rosie initiates the conversation and where the user initiates the conversation. The themes can be seen in figure 5.30 and 5.31.

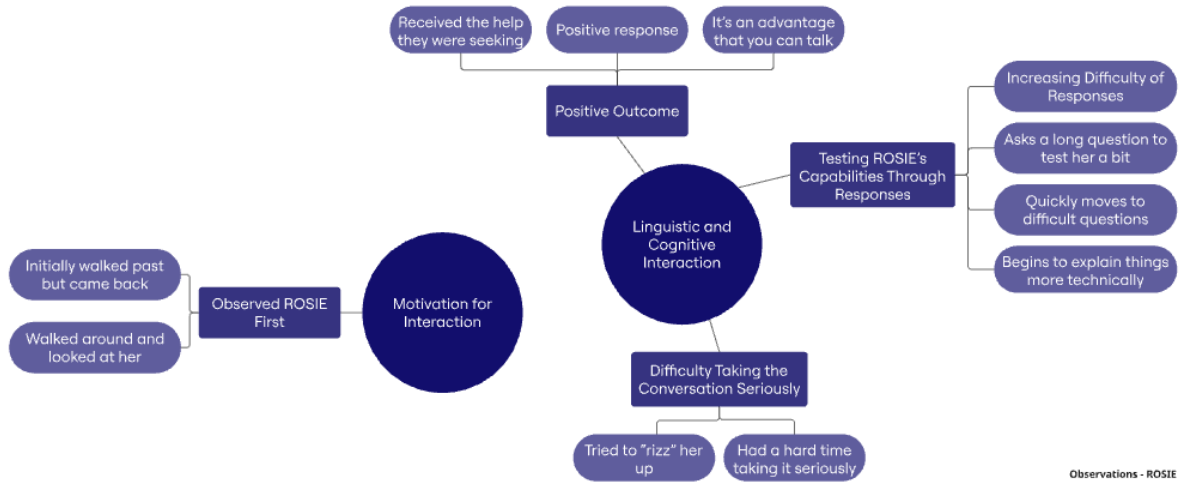

Figure 5.30: This figure shows the themes for the observations when Rosie initiates the conversation.

As figure 5.30 shows, the themes for the observations that has been noted when Rosie initiates the conversation are *Motivation for interaction* and *Linguistic and cognitive interaction*.

The subtheme for *Motivation for interaction* is *Observed Rosie first* and is based on participants who have previously walked past Rosie but came back at a later point. As well as participant who has observed her from a distance before walking up to her.

The subthemes for *Linguistic and cognitive interaction* were *Positive outcome, Testing Rosie's capabilities through responses*, and *Difficulty taking the conversation seriously*.

The subtheme *Positive outcome* is based on additional comments from participants who mention that they got the help that they were seeking and that it was an advantage that the interaction was based on speech.

The subtheme *Testing Rosie's capabilities through responses* is supported by observations of participants who are starting the interaction with short, simple, and non-technical answers and as the conversation develops they start to give long and technical answers. Other participants choose to quickly give complex response or ask a long question to test her capabilities.

The subtheme *Difficulty taking the conversation seriously* involves participants who are observed making fun prior to or during the interaction.

Figure 5.31: This figure shows the themes for the observations, when the participants initiate the conversation.

As seen in figure 5.31, the main themes for user-initiated conversations are *Situational dependency, Motivation for interaction*, and *Linguistic and cognitive interaction*.

The subtheme for *Situational dependency* is *Awkward situation*. This subtheme arises from participants who subsequently mentions that it is weird to talk to Rosie in the specific setting or participants who acts a bit goofy or silly because they are aware of the people around them.

The subthemes for *Motivation for interaction* are *Observed Rosie first, Familiar with the experiment*, and *Technological interest*. These participant have walked past Rosie to observe her and after some time returned to interact. Others have heard of the experiment and wanted to try it. Yet another group of participants are interested in the AI technology and wanted to test it.

For *Linguistic and cognitive interaction*, the subthemes are *Increasing difficulty of responses* and *Positive outcome*. These participants note that they had fun talking to Rosie and it was useful. Furthermore, they were observed to start the conversation with simple and short questions and as the conversation evolved, the difficulty increased.

5.5.5.1 Comparative Observations

In both conditions, participants were observed to increase the difficulty of their responses. Likely because they weren't sure of Rosie's capabilities and along the way noticed that she is capable of handling difficult and technical topics. This also affect the positive outcomes that the participants expressed in both conditions. The participants were able to get inputs on parts of their project which they struggled with and that was useful to them.

The situation in which the experiment is happening affects the interaction in both conditions. The participants think it's weird to have the conversation in public and this makes them act silly and gives them a hard time taking the conversation seriously.

The motivation for interaction differs slightly for the two conditions. Both when Rosie and the user initiates the conversation, some participants are observing Rosie and walks past her before they are willing to interact. But in the user-initiated conversation, the participation is also based on an interest in the technology and experiment. This suggests that the participants want to assess her first to build confidence unless they have a certain degree of curiosity or interest in interacting.

5.5.6 Other remarks

Throughout the study, it was observed that most of the participants interacted with Rosie while being surrounded by a group of friends. In fact 15/20 participant interacted with Rosie while being in a group. In the condition where Rosie initiated the conversation, 8/9 was in a group. In the user-initiated condition 7/11 was in a group.

The user-initiated condition had 11 participants. Out of those, three started the conversation by saying "Hello" and one said "Hey sweetie". Four participants started by asking the facilitator what they were supposed to do or whether it was possible to talk to Rosie. As they were told to try and say hello, those participants also initiated the conversation by saying "Hello". The two remaining participants asked the facilitators if there were any rules or if they had to scan the QR. They both end up saying "Hello".
This means that all participants initiate the conversation by saying hello. This is a common way

to start a conversation and at the same time, it gives participants the opportunity to align their expectations for the conversation based on the way Rosie reacts on their initiation.

Other observations during the experiment concerns some of the people who walked past Rosie and didn't interact. One of the fundamental observations was that people notice Rosie over the course of the week and take some time to warm up before joining a conversation. Also when Rosie started talking as people walked past, some of them gave comments such as "I don't want to talk to her, she's scary", "No", and "I don't dare to do that".

5.5.7 Summary of Findings: Experience and Perceived Interaction Quality

Participants generally described their interactions with the digital human as positive and manageable, though with varying levels of depth and engagement depending on who initiated the conversation.

When Rosie initiated, feedback focused on first impressions, tone, and flow. Participants often found the experience smooth or surprisingly natural but occasionally noted that her attitude felt overly enthusiastic or artificial. Rosie-initiated interactions were more likely to be judged affectively and based on surface-level cues.

In user-initiated interactions, participants gave more evaluative and reflective feedback, commenting on intelligence, helpfulness, and response nuance. Some praised Rosie's listening skills, while others pointed to shallow answers, technical issues, or rigid conversational flow. A few participants felt that the public setting made the interaction less comfortable.

Across both groups, key factors shaping the experience included perceived authenticity, AI tone, technical stability, and social context. The findings suggest that system-initiated conversations are judged more on initial tone and atmosphere, while user-initiated conversations invite deeper assessments of the AI's competence and social fit.

Most of the themes that has been uncovered through the thematic analysis is only backed up by little data and we don't see a tendency for the data to be fully saturated. This means that other themes may occur in future research within the topic of conversation initiation.

Chapter 6 - Discussion

The discussion will summarize the investigation made to answer the problem statement "What differences can be observed in user behavior and engagement when interactions with digital humans are initiated by different parties?"

This statement is explored via hypothesis testing, sentence analysis, qualitative observations, questionnaire data and exit interviews.

6.1 Hypothesis testing

The quantitative results for H_1, which examined whether AI initiation affects the probability of a conversation occurring, are inconclusive. The experiment was compromised by facilitators providing guidance to participants on how to interact with Rosie in cases where she did not initiate the conversation herself. As a result, there is insufficient evidence to reject $H0_1$, indicating that there isn't a difference in user behavior and engagement when conversation is initiated by different parties.

Similarly, the quantitative results for H_2, regarding whether AI initiation influences the subjective experience of the conversation, do not reveal any statistically significant differences for the four constructs tested: Engagement ($p = 0.271$), Intention to use ($p = 0.477$), Perceived Ease of Use ($p = 0.528$) or Perceived usefulness ($p = 0.602$). Therefore, there is also insufficient evidence to reject $H0_2$, indicating no difference in user experience when the conversation is initiated by different parties.

However, there were clear instances where participants in the Rosie initiates condition were stopped in their tracks and drawn into an interaction they might not have otherwise pursued. In one case, a group of students were walking with their backs to Rosie when one member glanced over their shoulder. At that moment, Rosie initiated the conversation, startling the group but ultimately starting a conversation.

Conversely, the opposite phenomenon was observed in the user initiates condition. In one instance, a participant stood in front of Rosie for 5–10 seconds, then gave up and walked away when nothing happened. It's possible the participant had previously seen Rosie initiate conversations with others

and expected her to make the first move again, leading them to assume she was malfunctioning when she remained passive.

6.2 Asking for help

However, there was a significantly higher rate of participants seeking assistance from the facilitators in the condition where Rosie did not initiate the conversation ($p = 0.012$). This suggests that users clearly interpret AI-initiated dialogue as a cue to engage, whereas in the absence of such initiation, they are uncertain about how to proceed. Whether users figure it out through experimentation in a real world setting is of course unknown but evidence from observing the use of Trothilde suggests that this is unlikely. Despite the presence of instructional signs ("Talk with me"), several participants appeared confused about how to engage, indicating that this phrasing may not sufficiently communicate the voice-based interaction modality, or perhaps users didn't read the signs. Conversely the qualitative data showed that other participants accredited the signs as the reason for their engagement.

6.3 Comparison of the four setups

The results in Figure 5.13 reveal a significant difference in the probability of success between setups 1 and 4 versus setups 2 and 3 with setup 1 and 4 having much more failures per hour (6-7.5 vs 0-1.5). Despite this there was a relatively consistent success rate per hour across all four setups. This discrepancy can be better understood by considering Rosie's placement and visibility in each setup. In Setup 2, for instance, Rosie is positioned between four pillars in a location that is rarely used as a walking path. Moreover, she is visible from a distance—something not clearly conveyed in Figure 5.5. As a result, no users encountered Rosie by chance in these setups; those who entered the square had likely already decided to interact with her. This may explain the variation in probability of success between setups 1,4 (ca. 11%) and 2,3 (ca. 50-100%) without a corresponding change in absolute successes per hour (ca. 0.7-1.2 across all four setups).

It suggests that the key factor may not be Rosie's behavior or location, but rather the characteristics of the individuals who choose to engage with her. This is both supported by the qualitative data and

the quantitative data. The quantitative data imply that only a subset of people are inclined to interact with Rosie, and that the frequency of such individuals passing by remains stable across locations. Notably the rate of people not wanting to interact could also be stable across locations though aren't captured when they can walk around the area in setup 2 and 3. It was observed that the participants who did interact seemed more socially confident or expressive than those who ignored Rosie. This is also supported by research by Bickmore and Cassell, 2005 who showed that extroverted users tend to prefer agents that initiate conversations and include small talk.
Therefore, future efforts should focus on identifying factors that increase willingness to engage across the whole population as well as enhancing the experience for those who are already motivated, or brave enough, to interact with digital humans. An investigation into the types of people who do interact with Rosie could be very interesting to understand which factors determine willingness to engage. A way to understand the scope of this issue could be to find various characteristics such as personality via the big five personality test and examine how that lines up with the population average.

6.4 Social context

The thematic analysis revealed that the social context in which Rosie was placed significantly influenced participants' willingness to engage. Several described the situation as awkward or strange, especially when other people were nearby. This aligns with psychological theories on social inhibition, which suggest that individuals in public settings often become more concerned with how they are perceived by others, particularly when they are uncertain about what is socially appropriate (Schlenker & Leary, 1982).

In our study, the cafeteria setting placed participants in a semi-public space where their behavior was visible to others. In such contexts, high social visibility combined with ambiguous expectations can lead individuals either to avoid interaction altogether or to engage in a playful, performative manner. This was echoed in several observations and participant remarks: many simply "thought it was weird to have the conversation in public," which made them act silly or exaggerated, often adopting humorous tones, asking absurd questions, or using exaggerated speech. This performative behavior suggests that some participants struggled to take the interaction seriously, perhaps as a

way to manage discomfort and reduce the risk of feeling judged by those watching. Drawing on Leary and Schlenker's model, this can be understood as a form of impression management: when unsure of how their actions will be interpreted, participants preemptively framed the interaction as non-serious (Schlenker & Leary, 1982). In doing so, they reduced the risk of negative evaluation and distanced themselves from the social consequences of engaging with an unfamiliar technology.

We also found distinct patterns in how different social constellations responded to Rosie. When Rosie initiated contact, it was primarily groups who stopped and interacted, often through shared curiosity or commentary. In fact, 8 out of 9 participants in the Rosie-initiated condition engaged with her as part of a group. These group interactions seemed to offer a kind of social buffer, making the situation feel safer. In contrast, the user-initiated condition included both individuals and groups, with 7 out of 11 participants engaging as part of a group, often with a clearer task orientation or purpose.

From a theoretical perspective, these patterns can be understood through the lens of social anxiety theory. When people are unsure of how to act, and when the stakes feel socially relevant as they often do in public, they are more likely to hesitate. However, in group settings, the perceived social risk is distributed, which can lower individual inhibitions and facilitate engagement (Schlenker & Leary, 1982).

6.5 General exit questions

The thematic analysis offers insight into participants' impressions, motivations, and evaluations of their interaction with the digital human Rosie. However, it is important to reflect on how the scope and phrasing of our exit interview questions have shaped the resulting themes. The questions were deliberately broad and open-ended—for instance: "Why did you stop and talk to her?", and "How did the conversation go overall?". This approach was intentional: our aim was not to evaluate the technical performance of the prototype in detail, but rather to explore the subjective experience of engaging in a spoken interaction with a digital human in a semi-public setting.

Given that the system is still a prototype with known limitations, such as slow response times, limited facial expressiveness, and occasional breakdowns, we chose not to direct participants'

attention toward specific system shortcomings. Instead, we sought to elicit reflections on the social and experiential dimensions of the interaction. As a result, the themes that emerged reflect broader impressions and affective responses, such as curiosity, awkwardness, or playfulness, rather than narrowly focused usability critiques.

That said, the open nature of our questions likely contributed to the emergence of overarching themes, rather than more granular distinctions. This trade-off was considered acceptable, given our exploratory aim and interest in contextualized user experience.

Looking back, however, one question we would have liked to include was: “What was it like to talk to an AI, and how did it feel to do so in a public space?” Such a prompt might have revealed more targeted insights into how users navigate the sociotechnical boundary between human-like interaction and artificial presence—particularly in a setting where others can observe. Given the central role of context in shaping interaction, this missed opportunity highlights the importance of attending not only to the interface itself, but also to the social dynamics that unfold around it.

6.6 Gamification

An interesting observation in the experiment is how participants often begin with simple or practical questions but gradually increase the complexity and difficulty of their interaction with Rosie. This stepwise escalation of demands on the system can be interpreted as a playful or challenging approach, where participants test Rosie’s boundaries and capabilities. It resembles the types of interactions identified by Ronagh Nikghalb and Cheng (2025) in their study of playful use of ChatGPT. In particular, the categories ’challenging’ and ’reflecting’ are highlighted, where users deliberately push the AI to explore its competence or “thinking.”

As Ronagh Nikghalb and Cheng (2025) describe, users engage with AI systems not merely to solve tasks, but to explore their limits, understand their capabilities, stumble upon serendipitous insights, and to have fun. Borrowing from Sicart, this process can be seen as a way of negotiating agency, where the user probes what the system can and cannot do, and how it positions itself in response. In this light, participants’ gradual increase in question complexity is not only a functional test, but also a social and interpretive act aimed at uncovering the contours of Rosie’s perceived intelligence.

In both cases, play is used as a means to understand the system's function and potential. Users are not only interested in what Rosie can respond to, but also in how she responds and whether she can keep up with complex or technical conversations. This demonstrates that playful interactions, even in a semi-formal or experimental context, can serve as a meaningful way for users to build trust, assess agency, and engage with the possibilities of the technology.

Importantly, this process of exploration and testing appears to depend on users having time and space to observe the system before engaging with it. Several participants lingered nearby or watched others before approaching, suggesting a need to "see it in action" first. This may explain why the digital humans at locations such as the job center or Krifa are less frequently used; contexts like these often lack the temporal or social affordances for such exploratory behavior. When there is no time to watch, no room to experiment, and no social permission to play, users may hesitate to initiate interaction at all.

6.7 Positioning Our Findings in Relation to Existing Literature

Our findings can be meaningfully contextualized within the existing body of research on embodied conversational agents (ECAs). While our quantitative analysis did not reveal a statistically significant difference in engagement between the two conditions, we did observe a tendency toward slightly higher engagement in the Rosie-initiated condition, particularly when excluding participants who asked facilitators for help. This aligns with findings from Van Pinxteren et al. (2023), who used the Elliot (1996) engagement scale and reported a mean of 4.65 (α = 0.90), with higher scores in socially oriented interaction styles. Although our results cannot confirm this relationship with statistical certainty, the tendency towards a higher score in the AI-initiated condition suggests that this condition may be perceived by users as a socially engaging cue - potentially facilitating interaction in more spontaneous, real-world contexts.

In terms of intention to use, our TAM-based measurements reflect a somewhat more cautious user response. Van Pinxteren et al. (2023) also included this construct using a four-item scale adapted from Taylor and Baker (1994), and reported a mean of 4.38 (α = 0.90). Their structured scenario and socially framed agents produced moderate-to-high intentions to use the technology. In contrast, our more naturalistic cafeteria setup, where users encountered Rosie spontaneously and

without instruction, may have been what caused lower intention scores for this study. This difference suggests that contextual framing and expectation shape how users evaluate digital systems.

Lastly, a major contribution of our study is its ecological validity. Unlike the majority of previous research from table 3.3, which has been conducted in laboratories or with scripted scenarios, our experiment was set in a university cafeteria, where users encountered the digital agent spontaneously. This setting allowed us to capture authentic reactions and decision-making processes, reflecting real-world complexities which are often missing from more controlled studies.

6.8 Voice interaction with digital humans

The research focused on a single feature which was conversation initiation and showed that this does impact the user experience by providing a strong enough signifier that users understood that they could, or at least felt comfortable enough to try, talking to the digital human thereby starting the voice interaction. From here the conversation got more interesting to observe as Rosie could show the conversational capabilities of a LLM leading to more nuanced conversations delving into technical topics and complex reasoning. Most likely they will expect this level of conversation and inference when interacting with other digital humans which would cement digital human interaction as a viable interface between users and customers.

Chapter 7 - Conclusion

This master's thesis explored the understanding, development, and testing of digital humans, with a particular focus on how conversational initiation influences user behavior. A review of the existing literature revealed that while many aspects of human–AI interaction have been studied, the specific role of AI-initiated dialogue remains underexplored. This gap informed the core research objective: *What differences can be observed in user behavior and engagement when interactions with digital humans are initiated by different parties?*

To investigate this, a digital human named Rosie was developed. Rosie was designed to operate in two distinct modes: either initiating conversations proactively or remaining silent until approached by a user. Two hypotheses were formulated: H1 tested whether AI initiation increases the probability of users engaging in conversation, and H2 examined whether initiation affects users' subjective experience of the interaction.

The public field experiment yielded no statistically significant difference in the overall rate of engagement between the two conditions, leading to the conclusion that H1 could not be supported based on the data collected. However, an important secondary finding emerged: users were significantly more likely to ask facilitators for help when Rosie did not initiate the conversation. This difference was statistically significant ($p = 0.012$), suggesting that users interpret AI initiation as a strong signifier for voice interaction and when the AI does not initiate, users can become confused.

Moreover, because facilitators provided users with guidance on how to start the interaction when asked, the validity of the results concerning H1 was further complicated. This assistance may have diminished the observable effect of initiation on engagement, introducing a potential confounding factor in the study design.

The second hypothesis (H2), which concerned potential differences in the subjective quality of the interaction, was investigated through observational notes, a questionnaire measuring the constructs "Engagement" and "Intention to Use," as well as exit interviews. The questionnaire results revealed no statistically significant differences between the two conditions, leading to the conclusion that H2 could not be supported based on the quantitative data.

However, a thematic analysis of the qualitative data revealed a clear distinction in how users perceived Rosie under the two conditions. Users who had to initiate the conversation themselves

often felt uncertain about how to begin and were more likely to reflect on the quality of the conversation and attribute greater cognitive abilities to the digital human. In contrast, users who experienced Rosie initiating the interaction tended to focus on Rosie's physical attributes, such as her graphical fidelity.

In summary, the results suggest that AI-initiated interaction serves as a strong cue for users, indicating that voice interaction is possible and expected. When this cue is absent, users are often unsure how to proceed. Therefore, enabling digital humans to initiate conversation may improve the quality and naturalness of interactions in real-world settings.

Chapter 8 - Future work

For future purposes Rosie could be used as a research platform as she is open source and interfaces with popular tools which automatically will follow technological progression via OpenAI's evolving API. The more interesting contribution of this research is the methodological considerations developed through the iterations of the experiment. The research presented a lot of interesting challenges such as participant personality, facilitator presence and placement in the context of public spaces. Controlling and understanding the impact of these factors should be done in order to design a more engaging and easily understandable digital human receptions.

In addition, future research could investigate interactions in other contexts, such as private or semi-private settings, where the gaze of the public does not play the same role. These environments may foster greater willingness to explore and engage playfully with AI systems. Furthermore, longitudinal user scenarios combined with follow-up interviews could provide a deeper understanding of how such AI agents are perceived over time, and which factors contribute to sustained trust and acceptance.

If this research is to serve as a foundation for further studies, several areas should be explored to enable more automated and scalable experimentation. Most notably, Rosie would need to autonomously log user activity, including detecting when users approach, initiate interaction, and when a conversation has ended. This kind of automated sensing and logging is essential for minimizing facilitator involvement and increasing experimental validity in real-world deployments.

Bibliography

Abel, M., Kuz, S., Patel, H. J., Petruck, H., Schlick, C. M., Pellicano, A., & Binkofski, F. C. (2020). Gender effects in observation of robotic and humanoid actions. [Place: Switzerland Publisher: Frontiers Media S.A.]. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00797

Agresti, A. (2018). *Statistical methods for the social sciences* (Fifth edition, global edition). Pearson Education, Limited.

Ahn, J., Kim, J., & Sung, Y. (2021). AI-powered recommendations: The roles of perceived similarity and psychological distance on persuasion. [Place: United Kingdom Publisher: Taylor & Francis]. *International Journal of Advertising: The Review of Marketing Communications*, *40*(8), 1366–1384. https://doi.org/10.1080/02650487.2021.1982529

Akdim, K., Belanche, D., & Flavián, M. (2021). Attitudes toward service robots: Analyses of explicit and implicit attitudes based on anthropomorphism and construal level theory [Publisher: Emerald Publishing Limited]. *International Journal of Contemporary Hospitality Management*, *35*(8), 2816–2837. https://doi.org/10.1108/IJCHM-12-2020-1406

Alstrup, C. (2025). Resultater af tværkommunal borgerundersøgelse på beskæftigelsesområdet. https://www.kl.dk/beskaeftigelse-og-integration/beskaeftigelsesindsatsen/tvaerkommunal-borgerundersoegelse-paa-beskaeftigelsesomraadet

Bae, S., Jung, T., Cho, J., & Kwon, O. (2025). Effects of meta human characteristics on user acceptance: From the perspective of uncanny valley theory. *Behaviour and Information Technology*, *44*(4), 731–748.

Bailenson, J. N., & Yee, N. (2005). Digital Chameleons: Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *Psychological Science*, *16*(10), 814–819. https://doi.org/10.1111/j.1467-9280.2005.01619.x

Baron, N. S. (2015). Shall We Talk? Conversing With Humans and Robots. *The Information Society*, *31*(3), 257–264. https://doi.org/10.1080/01972243.2015.1020211

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures: *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Becker, C., Kopp, S., & Wachsmuth, I. (2007, November). Why Emotions should be Integrated into Conversational Agents. In T. Nishida (Ed.), *Conversational Informatics* (1st ed., pp. 49–67). Wiley. https://doi.org/10.1002/9780470512470.ch3

Ben Mimoun, M. S., & Poncin, I. (2015). A valued agent: How ECAs affect website customers' satisfaction and behaviors. *Journal of Retailing and Consumer Services*, *26*, 70–82. https://doi.org/10.1016/j.jretconser.2015.05.008

Benitez, J., Wyman, A. B., Carpinella, C. M., & Stroessner, S. J. (2017). The authority of appearance: How robot features influence trait inferences and evaluative responses [ISSN: 1944-9437]. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 397–404. https://doi.org/10.1109/ROMAN.2017.8172333

Bergmann, K., Eyssel, F., & Kopp, S. (2012). A Second Chance to Make a First Impression? How Appearance and Nonverbal Behavior Affect Perceived Warmth and Competence of Virtual Agents over Time [Series Title: Lecture Notes in Computer Science]. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, Y. Nakano, M. Neff, A. Paiva, & M. Walker (Eds.), *Intelligent Virtual Agents* (pp. 126–138, Vol. 7502). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33197-8_13

Biancardi, B., Dermouche, S., & Pelachaud, C. (2021). Adaptation Mechanisms in Human-Agent Interaction: Effects on User's Impressions and Engagement.

Bickmore, T., & Cassell, J. (2005). Social Dialongue with Embodied Conversational Agents [Series Title: Text, Speech and Language Technology]. In N. Ide, J. Véronis, H. Baayen, K. W. Church, J. Klavans, D. T. Barnard, D. Tufis, J. Llisterri, S. Johansson, J. Mariani, J. C. J. Van Kuppevelt, L. Dybkjær, & N. O. Bernsen (Eds.), *Advances in Natural Multimodal Dialogue Systems* (pp. 23–54, Vol. 30). Springer Netherlands. https://doi.org/10.1007/1-4020-3933-6_2

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. [Place: Germany Publisher: Springer]. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cerekovic, A., Aran, O., & Gatica-Perez, D. (2017). Rapport with Virtual Agents: What Do Human Social Cues and Personality Explain? *IEEE Transactions on Affective Computing*, *8*(3), 382–395. https://doi.org/10.1109/TAFFC.2016.2545650

Čereković, A., Huang, H.-H., Furukawa, T., Yamaoka, Y., Pandžić, I., Nishida, T., & Nakano, Y. (2009). Implementing a multi-user tour guide system with an embodied conversational agent. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5820 LNCS*, 7–18. https://doi.org/10.1007/978-3-642-04875-3_7

Chateau, N., Maffiolo, V., Pican, N., & Mersiol, M. (2005). The Effect of Embodied Conversational Agents' Speech Quality on Users' Attention and Emotion [Series Title: Lecture Notes in Computer Science]. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. Tao, T. Tan, & R. W. Picard (Eds.), *Affective Computing and Intelligent Interaction* (pp. 652–659, Vol. 3784). Springer Berlin Heidelberg. https://doi.org/10.1007/11573548_84

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, *90*, 315–330. https://doi.org/10.1016/j.chb.2018.08.048

Coghlan, S. (2024). Anthropomorphizing machines: Reality or popular myth? [Place: Germany Publisher: Springer]. *Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science*, *34*(3). https://doi.org/10.1007/s11023-024-09686-w

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Diel, A., Weigelt, S., & Macdorman, K. F. (2022). A Meta-analysis of the Uncanny Valley's Independent and Dependent Variables. *ACM Transactions on Human-Robot Interaction*, *11*(1), 1–33. https://doi.org/10.1145/3470742

Dirk Heylen, Mariët Theune, Rieks op den Akker, & Anton Nijholt. (2009). Social agents: The first generations. https://doi.org/10.1109/ACII.2009.5349476

Ehret, J., Bönsch, A., Aspöck, L., Röhr, C. T., Baumann, S., Grice, M., Fels, J., & Kuhlen, T. W. (2021). Do prosody and embodiment influence the perceived naturalness of conversational agents' speech? [Place: US Publisher: ACM - Association for Computing Machinery]. *ACM Transactions on Applied Perception*, *18*(4). https://doi.org/10.1145/3486580

Elliot, A. J. (1996). Approach and Avoidance Achievement Goals and Intrinsic Motivation: A Mediational Analysis. *Journal of Personality and Social Psychology*, *Vol. 70*(No.3), 461–475.

Erengin, T., Briker, R., & de Jong, S. B. (2025). You, me, and the ai: The role of third-party human teammates for trust formation toward ai teammates. [Place: US Publisher: John Wiley & Sons]. *Journal of Organizational Behavior*, No Pagination Specified–No Pagination Specified. https://doi.org/10.1002/job.2857

Esposito, A., & Jain, L. C. (Eds.). (2016). *Toward Robotic Socially Believable Behaving Systems - Volume II* (Vol. 106). Springer International Publishing. https://doi.org/10.1007/978-3-319-31053-4

Fakhimi, A., Garry, T., & Biggemann, S. (2023). The effects of anthropomorphised virtual conversational assistants on consumer engagement and trust during service encounters. [Place: US Publisher: Sage Publications]. *Australasian Marketing Journal (AMJ)*, *31*(4), 314–324. https://doi.org/10.1177/14413582231181140

Ferdinand Engländer. (2014, June). The Uncanny Valley - Animator Island [Section: Acting]. Retrieved February 5, 2025, from https://www.animatorisland.com/the-uncanny-valley/

Foster, M. E. (2007). Enhancing Human-Computer Interaction with Embodied Conversational Agents [Series Title: Lecture Notes in Computer Science]. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, & C. Stephanidis (Eds.), *Universal Access in Human-Computer Interaction. Ambient Interaction* (pp. 828–837, Vol. 4555). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73281-5_91

Geiselmann, R., Tsourgianni, A., Deroy, O., & Harris, L. T. (2023). Interacting with agents without a mind: The case for artificial agents. [Place: Netherlands Publisher: Elsevier Science]. *Current Opinion in Behavioral Sciences*, *51*. https://doi.org/10.1016/j.cobeha.2023.101282

Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M., Van Der Werf, R. J., & Morency, L.-P. (2007). Can Virtual Humans Be More Engaging Than Real Ones? [Series Title: Lecture Notes in Computer Science]. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, & J. A. Jacko (Eds.), *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments* (pp. 286–297, Vol. 4552). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73110-8_30

Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J. K., & Robles, E. (2009). Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies*, *67*(10), 842–849. https://doi.org/10.1016/j.ijhcs.2009.07.001

Guadagno, R. E., Swinth, K. R., & Blascovich, J. (2011). Social evaluations of embodied agents and avatars. [Place: Netherlands Publisher: Elsevier Science]. *Computers in Human Behavior*, *27*(6), 2380–2385. https://doi.org/10.1016/j.chb.2011.07.017

Jeppe Givskud. (2025a, April). The first functional interaction with iteration 1 of Rosie - Interaction is in danish. Retrieved May 27, 2025, from https://www.youtube.com/watch?v=gKidbRIFDds

Jeppe Givskud. (2025b, May). Experiment 1 - Rosie initiates conversation. Retrieved May 27, 2025, from https://www.youtube.com/watch?v=uAh9NrMyKeE

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Intelligent Virtual Agents* (pp. 329–343). Springer. https://doi.org/10.1007/11550617_28

Kramer, N. (2005). Social communicative effects of a virtual program guide, 442–453.

Krämer, N. (2008). Social effects of virtual assistants. A review of empirical results with regard to communication. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5208 LNAI*, 507–508. https://doi.org/10.1007/978-3-540-85483-8_63

Lee, M. K., Kiesler, S., & Forlizzi, J. (2010). Receptionist or information kiosk: How do people talk with a robot? *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 31–40. https://doi.org/10.1145/1718918.1718927

Lombard, M. (2011). Measuring Telepresence: The Validity of The Temple Presence Inventory (TPI) in a Gaming Context.

Loveys, K., Hiko, C., Sagar, M., Zhang, X., & Broadbent, E. (2022). "I felt her company": A qualitative study on factors affecting closeness and emotional support seeking with an embodied conversational agent. *International Journal of Human-Computer Studies*, *160*, 102771. https://doi.org/10.1016/j.ijhcs.2021.102771

Loveys, K., Sagar, M., & Broadbent, E. (2020). The Effect of Multimodal Emotional Expression on Responses to a Digital Human during a Self-Disclosure Conversation: A Computational Analysis of User Language. *Journal of Medical Systems*, *44*(9), 143. https://doi.org/10.1007/s10916-020-01624-4

Loveys, K., Sebaratnam, G., Sagar, M., & Broadbent, E. (2020). The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review. *International Journal of Social Robotics*, *12*(6), 1293–1312. https://doi.org/10.1007/s12369-020-00680-7

Martin, A., & Jenkins, K. (2024). Speaking your language: The psychological impact of dialect integration in artificial intelligence systems. [Place: Netherlands Publisher: Elsevier Science]. *Current Opinion in Psychology*, *58*. https://doi.org/10.1016/j.copsyc.2024.101840

Network, M. (2024, February). The Evolution of AI-Powered Personal Assistants: A Comprehensive Guide to Siri, Alexa, and Google... Retrieved April 3, 2025, from https://megasisnetwork.medium.com/the-evolution-of-ai-powered-personal-assistants-a-comprehensive-guide-to-siri-alexa-and-google-f2227172051e

Norman, D. A. (2002). *The design of everyday things*. Basic Books, Inc.

Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., & Peters, C. (2020). Engagement in Human-Agent Interaction: An Overview. *Frontiers in Robotics and AI*, *7*, 92. https://doi.org/10.3389/frobt.2020.00092

OpenAI. (2025a). ChatGPT. Retrieved May 27, 2025, from https://chatgpt.com

OpenAI. (2025b). Quickstart - OpenAI Agents SDK. Retrieved May 27, 2025, from https://openai.github.io/openai-agents-python/voice/quickstart/

OpenAI. (2025c, June). Openai/whisper [original-date: 2022-09-16T20:02:54Z]. Retrieved June 2, 2025, from https://github.com/openai/whisper

Park, S. Y. (2009). An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning.

Pickard, M. D., Burgoon, J. K., & Derrick, D. C. (2014). Toward an Objective Linguistic-Based Measure of Perceived Embodied Conversational Agent Power and Likeability. *International Journal of Human-Computer Interaction*, *30*(6), 495–516. https://doi.org/10.1080/10447318.2014.888504

Potdevin, D., Clavel, C., & Sabouret, N. (2021). A virtual tourist counselor expressing intimacy behaviors: A new perspective to create emotion in visitors and offer them a better user experience? *International Journal of Human-Computer Studies*, *150*, 102612. https://doi.org/10.1016/j.ijhcs.2021.102612

Rebecca Hansen. (2025). *Praktik Ved NTT DATA buisness solution* (Master).

Robinson, S., Traum, D., Ittycheriah, M., & Henderer, J. (2008). What would you ask a Conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting.

Ronagh Nikghalb, M., & Cheng, J. (2025). Interrogating AI: Characterizing Emergent Playful Interactions with ChatGPT. *Proceedings of the ACM on Human-Computer Interaction*, *9*(2), 1–23. https://doi.org/10.1145/3711015

Roso, N., Vlahović, S., Markuš, N., & Sužnjević, M. (2024). Exploring Conversations with AI NPCs: The Impact of Token Latency on QoE and Player Experience in a Text-Based Game. *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)*, 258–261. https://doi.org/10.1109/QoMEX61742.2024.10598251

Schlenker, B. R., & Leary, M. R. (1982). Social Anxiety and Self-Presentation: A Conceptualization and Model.

Sharma, M., & Vemuri, K. (2022). Accepting Human-like Avatars in Social and Professional Roles. *ACM Transactions on Human-Robot Interaction*, *11*(3), 1–19. https://doi.org/10.1145/3526026

Smith, C., Crook, N., Boye, J., Charlton, D., Dobnik, S., Pizzi, D., Cavazza, M., Pulman, S., de la Camara, R. S., & Turunen, M. (2010). Interaction Strategies for an Affective Conversational Agent. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Intelligent Virtual Agents* (pp. 301–314). Springer. https://doi.org/10.1007/978-3-642-15892-6_31

Stein, J.-P., Appel, M., Jost, A., & Ohler, P. (2020). Matter over mind? How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both. *International Journal of Human-Computer Studies*, *142*, 102463. https://doi.org/10.1016/j.ijhcs.2020.102463

Sung, E. (, Han, D.-I. D., Choi, Y. K., Gillespie, B., Couperus, A., & Koppert, M. (2023). Augmented digital human vs. human agents in storytelling marketing: Exploratory electroencephalography and experimental studies. [Place: US Publisher: John Wiley & Sons]. *Psychology & Marketing*, *40*(11), 2428–2446. https://doi.org/10.1002/mar.21898

Tabalba, R., Lee, C. J., Tran, G., Kirshenbaum, N., & Leigh, J. (2024, September). ArticulatePro: A Comparative Study on a Proactive and Non-Proactive Assistant in a Climate Data Exploration Task [arXiv:2409.10797 [cs]]. https://doi.org/10.48550/arXiv.2409.10797

Taylor, S. A., & Baker, T. L. (1994). An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions. *Journal of Retailing*, *70*(2), 163–178. https://doi.org/10.1016/0022-4359(94)90013-2

Ter Stal, S., Tabak, M., Op Den Akker, H., Beinema, T., & Hermens, H. (2020). Who Do You Prefer? The Effect of Age, Gender and Role on Users' First Impressions of Embodied Conversational Agents in eHealth. *International Journal of Human–Computer Interaction*, *36*(9), 881–892. https://doi.org/10.1080/10447318.2019.1699744

Van Pinxteren, M. M. E., Pluymaekers, M., Lemmink, J., & Krispin, A. (2023). Effects of communication style on relational outcomes in interactions between customers and embodied conversational agents. *Psychology & Marketing*, *40*(5), 938–953. https://doi.org/10.1002/mar.21792

Van Pinxteren, M. M., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like communication in conversational agents: A literature review and research agenda. *Journal of Service Management*, *31*(2), 203–225. https://doi.org/10.1108/JOSM-06-2019-0175

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). (PDF) Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. *ResearchGate*. Retrieved April 3, 2025, from https://www.researchgate.net/publication/259440652_Reconstructing_the_Giant_On_the_Importance_of_Rigour_in_Documenting_the_Literature_Search_Process

Wagner, K., Nimmermann, F., & Schramm-Klein, H. (2019). Is It Human? The Role of Anthropomorphism as a Driver for the Successful Acceptance of Digital Voice Assistants.

Watson, J., Valsesia, F., & Segal, S. (2024). Assessing AI receptivity through a persuasion knowledge lens. [Place: Netherlands Publisher: Elsevier Science]. *Current Opinion in Psychology*, *58*. https://doi.org/10.1016/j.copsyc.2024.101834

Xavier, E. K., Putra, J., & Anita, T. L. (2023). Customer satisfaction from the self – service kiosks' UI/UX and the customer continuance intention to use (T. Mursitama, Noerlina, D. Utama, & S. Abrori, Eds.). *E3S Web of Conferences*, *426*, 01083. https://doi.org/10.1051/e3sconf/202342601083

Zhang, Y. (, Tan, W., & Lee, E.-J. (2024). Consumers' responses to personalized service from medical artificial intelligence and human doctors. [Place: US Publisher: John Wiley & Sons]. *Psychology & Marketing*, *41*(1), 118–133. https://doi.org/10.1002/mar.21911

Zhang, Z., Guo, F., Fang, C., & Chen, J. (2025). Let me hold your hand: Effects of anthropomorphism and touch behavior on self-disclosure intention, attachment, and cerebral activity towards ai mental health counselors. [Place: United Kingdom Publisher: Taylor & Francis]. *International Journal of Human-Computer Interaction*, No Pagination Specified–No Pagination Specified. https://doi.org/10.1080/10447318.2024.2446502

Appendix A | Thematic Analysis of Semi-Structured Interview

In order to better understand the situated use and reception of the digital human Sophia (formerly known as Aida), a semi-structured interview was conducted with a staff member at a site where the technology has been in use for approximately two years. The conversation focused on both user interaction and organizational adaptation. While limited to a single informant, the interview offered a detailed and experience-based account of how Sophia is used in everyday practice, how users typically respond, and how staff have adapted their routines accordingly. The full interview transcript is available in Attachment "Krifa Interview".

A Methodological Approach

Instead of full transcription, the audio recording of the interview was listened to carefully, and notes taken during the meeting were reviewed and supplemented where necessary. This ensured that key points and examples were accurately captured, even if not all details were documented verbatim. Particular attention was paid to recurring challenges, practical examples, and broader reflections related to the use of the digital receptionist Sophia. While the findings are based on a single staff member's perspective, the insights provide valuable indications of both technical and social dynamics involved in everyday use of the technology.

Attention was given to both concrete examples and broader reflections on implementation and use. Through this process, a number of recurring issues and observations were identified ranging from technical performance and user behavior to internal workflows and cultural perceptions. Although the findings are based on a single perspective, the informant's practical experience offers valuable real-world insight into both the benefits and challenges of integrating a digital receptionist in a high-traffic front-desk environment.

A.1 Theme 1: Visibility and Awareness

One of the most immediate barriers to interaction is the lack of visibility. Users often do not notice Sophia or do not realize that she is intended to replace the receptionist. This results in users bypassing her entirely and either approaching office staff or seeking help elsewhere. Signage and placement could play a key role in this issue.

"It's not always that people notice Sophia... then they walk up to the door and ask in the office."

The participant also described how users rely on observing others in the space to understand what to do, suggesting that peer behavior influences initial interaction patterns.

A.2 Theme 2: User Behavior and Habits

The interview revealed clear trends in how users approach the digital receptionist. Most users prefer to interact with the touchscreen, especially when time is limited or the technology appears slow to respond. Habituation and expectations of speed and ease from other digital services shape these behaviors.

"Most people end up using the touchscreen, because she's sometimes a bit slow."

Social mimicry was also observed: if one user approached the screen or bypassed Sophia, others tended to follow that pattern.

"People just do what they see others doing."

A.3 Theme 3: Technical Limitations and Frustration Points

Technical limitations were a recurring concern. Users frequently experienced delayed responses, misunderstood inputs, or situations where Sophia would abruptly reset the interaction. Some users even resorted to raising their voice to get a response. The system also lacked clear feedback, making it unclear whether it was processing a request or had stopped listening.

"Sometimes she just doesn't respond at all, and people start yelling." "When she doesn't work, we have to do the receptionist's job ourselves."

In some cases, physical signs such as “Sophia is on sick leave” had to be placed to communicate system downtime, illustrating how technical issues impact trust and continuity in service delivery. This also shows how krifa is, at least on the surface, anthropomorphizing Sophia by giving her human attributes such as sick days.

A.4 Theme 4: Acceptance and User Attitudes

Attitudes toward Sophia varied significantly depending on demographic and situational factors. Business clients were described as curious and generally receptive, whereas union members and other walk-in visitors often expressed skepticism. Openness to the technology was said to correlate with age, familiarity with digital tools, and emotional readiness.

“Business people often find her very interesting… members want to be met by a human.” “Young people are usually more curious.”

Moreover, users who initially resisted the technology were sometimes able to adjust over time and become more comfortable with using it.

A.5 Theme 5: Perceived Value and Organizational Impact

From an organizational perspective, the implementation of Sophia has led to increased efficiency, particularly in handling routine queries and check-ins. While there was some skepticism among staff in the early stages, Sophia is now generally viewed as a supportive tool that allows employees to focus on more complex or sensitive tasks.

“It’s just easier that she’s there.” “It really helps streamline things, especially the small stuff.”

Interestingly, the participant noted that some conversations are easier to initiate with a digital receptionist—particularly when disclosing personal or sensitive information.

“Sometimes it’s actually easier to just say your name to Sophia than explain everything to a real person.”

A.6 Theme 6: Human Support and Social Context

Despite the presence of Sophia, human staff continue to play an essential role. Staff members often monitor users discreetly and step in before confusion escalates. This "service instinct" ensures that users do not become stuck or frustrated. The participant emphasized that human presence remains vital, particularly in supporting users who struggle with digital interfaces.

"You can tell when someone is struggling – otherwise we just let them be."

The need for a "backup" reception function was also acknowledged, especially for unplanned visits or users with additional needs. The organization has adapted its physical layout and staff practices accordingly.

B Conclusion

This thematic analysis provides a multifaceted understanding of how a digital human is received, used, and supported within an organizational context. Visibility, speed, and system responsiveness are key factors that shape user behavior. While some users embrace the digital solution, others require time, support, or a fallback option. Human presence remains crucial for ensuring inclusivity and maintaining service quality. These findings underline the importance of integrating digital solutions into the broader social and spatial environment, and they point toward a need for design interventions that enhance clarity, responsiveness, and trust.

Appendix B | Variables in Articles from Literature Review

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Wilingness to return	zhang 2024	Personalization	Better with personalization but worse that human agents		
Persuation	watson 2024	User Personality	Some individuals are less persuative and some people want to do a task themselves		
Persuation	watson 2024	Perceived capabilities	More believed capabilities means more acceptance and trust	e.g.,ability to capture data,predict outcomes,produce,and interact with humans) are informed by sociological contextual elements Puntoni S,Reczek RW,Giesler M,Botti S: Consumers and artificial intelligence: an experiential perspective. J Mark 2021,85:131-151,https://doi.org/10.1177/0022242920953847.	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Persuation	watson 2024	Symbolic value of task	More important and symbolic interaction yiels less persuation	In the medical domain,for instance,individuals perceive medical AI as less capable of accounting for contextual factors than human medical providers,resulting in a decreased likelihood of engaging with AI	Kan vi teste det i kontektsen af jobcenteret. Så måske accepterer de ligegyldige ting som parkering blankt men ting som penge mindre blankt?
Effective communication	Sung et. al. 2023	Humanlike appearance	Better conviction when digital look more like humans than machines		watson2024 siger noget lidt andet
Attitudes towards service robots	Akdim et. al. 2023	Humanlike appearance	The realistic robots lead to both explicit and implicit negative attitudes	Robots with lower humanlikeness levels generate relatively more positive attitudes and are accepted to nearly the same extent as human employees in hospitality and tourism contexts	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Rapport	vanpinxteren 2023	Hedonic vs utilitarian	Hedonic favors social style and adaptive style		
Satisfaction and intention to use	vanpinxteren 2023	Rapport	Rapport is a good proxy for satisfaction and intention to use		
Engagement and trust	bickmore 2005	Embodiement,Task VS social	Voice er god til introverte og opgaveorienteret		
Customers' perception of empathy	vanpinxteren 2020	Pronoun choice	customer service agents using the pronoun "I" (firstperson singular) instead of "you" (second-person singular) or "we" (first-person plural) are perceived as more empathetic by customers	(Packard et al.,2014)	
sensitiveness and mechanicalness	vanpinxteren 2020	apologizing for their mistakes	has revealed that computers apologizing for their mistakes are perceived as more sensitive and less mechanical	Tzeng (2004)	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Psychological distance	ahn 2021	perceived similarity of recommendation agents	More similar = Lesser psychologysical distance	Human agent vurderes som mere psykologisk tæt end AI agent	Den er ikke embodied og de interagerer ikke med den men lytter kun til dens udsagn
The use of intelligent agents	benmimoun 2015	Search time and costs (how much time and resources are required for information retrieval).	observes that using intelligent agents to search for information reduces search time and costs	Chang (2010)	
Trust	benmimoun 2015	Small talk	Small talk was shown to improve perceived trust,knowledge of the user,and success of the interaction for extroverted users,while taskoriented talk received higher ratings from introverted users		

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
warmth	benmimoun 2015	friendly/neutral language	In another study,ECAs with friendly language were rated higher in warmth than those with neu tral language	Griffiths S,Eyssel FA,Philippsen A,Pietsch C,Wachsmuth S (2015)	
flow	benmimoun 2015	voice	Another study found that an artificial texttospeech voice was perceived to have better flow than textonly communication	Results suggest that a human like voice may be the most preferable option,however in the absence of a humanlike voice,a texttospeech voice is preferable to text only	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
trust	benmimoun 2015	personality	In this study,an extroverted ECA (faster speech rate,larger pitch range,frequent smiles,and expansive head ges ture) was rated as less trustworthy than an introverted ECA (slower speech rate,calm vocal tone,neutral facial expres sion,and low head animation) for extroverted participants. There were no significant differences in willingness to trust an extroverted or introverted ECA for introverted participants.		

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
rapport	chattaraman 2019	reciprocity principle	In conversing with a virtual agent,users may demonstrate the reciprocity principle by taking turns in conversing with the agent,pausing with the agent in conversation,and offering lengthy responses to the agent (Cerekovic,Aran,and Gatica-Perez,2017). Cerekovic et al. (2017) argued that these reciprocal behaviors are more evident when the user was more extraverted and agreeable by nature,but in general,the reciprocation strengthened the user's perception of rapport with the agent.	(Cerekovic,Aran,and GaticaPerez,2017).	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Task vs Social	chattaraman 2019	ældre bruger	In fact,Velestianos (2009,2012) found in two studies that nontask contexts of agents were distracting to learners. While examining socialversus taskoriented pedagogical agents,Veletsianos (2012) found qualitative evidence indicating that ontask conversation assisted participants in staying focused on a task at hand,and nontask (social) commentary of agents made tasks distracting,perhaps breaching users' cognitive load.	Velestianos (2009,2012)	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Cognitive load,measured through users' speech disfluency rates (i.e.,the number of disfluencies per 100 words)	chattaraman 2019	Type of dialog (social vs. taskoriented dialog).	Studiet fandt,at opgaveorienteret dialog førte til en højere kognitiv belastning (flere taleuflydigheder) sammenlignet med social dialog. Dette betyder,at når brugerne deltager i strukturerede,målrettede samtaler (som at diskutere specifikke ejendomskøb),oplever de en større mental belastning end ved mere uformelle,sociale samtaler.	Bickmore and Cassell (2005)	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Brugeradfærd (ændringer i brugeradfærd som respons på fejlmeddelelser).	bickmore 2006	Humanness	Resnick and Lammers showed that in order to change user behaviour via corrective error messages,the messages should have different degrees of “humanness” depending on whether the user has high or low selfesteem (“computerese” messages should be used with low selfesteem users,while “humanlike” messages should be used with high esteem users)	Resnick and Lammers,1985	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Subjective assessments of the agent	bickmore 2006	INITIATION (Active vs. Passive)	Observation of the videotaped data made it clear that some subjects took the initiative in the conversation,while others allowed REA to lead. Unfortu nately,REA is not yet able to deal with userinitiated talk,and so user initiative often led to REA interrupting the speaker. To assess the effect of this phenom enon,we therefore divided subjects into PASSIVE (below the mean on number of userinitiated utterances) and ACTIVE (above the mean on number of user initiated utterances). To our surprise,these measures turned out to be indepen dent of introversion/extroversion (Pearson r=0.042),and to not be predicted by these latter variables		

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Antropomorfism	blut 2021	Competence	Maybe competence facilitates antropomorfism but also not if the user is very technical	It seems that the more people are capable of using a robot,the lower their anthropomorphic tendency,because there is no need to facilitate the interaction by humanizing the robot." (Blut et al.,2021,p. 636)	
Antropomorfism	blut 2021	Need for interaction	More need for interaction = mere antropomorphism	Lonely people antropomorphise more	
Antropomorfism	blut 2021	Negative attitudes toward robots in daily life (NARS)	Mere NARS burde betyde mere antropomorfisme		
Antropomorfism	blut 2021	Age	Elderly and children antropomorphise more		
Antropomorfism	blut 2021	Physical human features	More human features = more anthropomorfism	head,face,and body	

Dependent Variable	**Sources**	**Independent Variable**	**Conclusion**	**Explanation**	**Tanker**
Antropomorfism	blut 2021	Nonphysical human features	More humanlike nonphysical features = More anthropomorfism.	eye contact,use gestures,move,talk,emotionality and personality	
Perceived intelligence	blut 2021	Antropomorfism	More Antropomorfism means higher perceived intelligence (also higher expectations for this reason)		
Percieved safety	blut 2021	Anthropomorfism	the more a customer perceives a service robot as human like,the safer the service experience appears	anthropomorphism can facilitate perceptions of safety by increasing the sense of the predictability and controllability of the nonhuman agent during interactions,thereby reducing feelings of risk and danger.	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Social prestence	blut 2021	Anthropomorfism	More anthropomorphic = More social prestence	Robots that are perceived as more human like can provide customers with a stronger social presence,thereby enriching social interaction.	
Usefulness	blut 2021	Anthropomorfism	More anthropomorfism more usefull		
Negative affect (negative feelings)	blut 2021	Anthropomorfism	Due to uncanny valley more Anthropomorfism can result in bad feelings		
Positive affect (positive feelings)	blut 2021	Anthropomorfism	Anthropomorfism can enhance positve affect	It seems that anthropomorphism can elicit both positive and negative emotions to ward a robot,with opposite effects on customer use intention,making this relationship complex	

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Rapport	blut 2021	Anthropomorfism	Anthropomorphism facilitates human-robot rapport,making it easier,more desirable,and more meaningful		
Satisfaction	blut 2021	Anthropomorfism	Generally expected to be better but due to higher expectations it can backfire if the robot is bad (trothilde)		
Trust	blut 2021	Anthropomorfism	People trust anthropomorphic products more in general		

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker
Natural speech	Social communicative effects of a virtual program guide	Humanlike features	When confronted with a humanlike character participants felt free to address it by natural speech. The use of personalized requests and words such as 'you' also increased	Text interface,GUI and ECA are tested in a setup where the subjects are asked to complete the same task on each of the interfaces using speach. The results show that the more humanlike the interface looks,the more natural language is used.	Another practical problem arises when during the systems processing time the face does not show an immediate reaction of understanding or misunderstanding or at least indicates that the user has to wait for further feedback. Our result of people frequently repeating their input when interacting with an agent points to the danger of presenting a face that sometimes does not react immediately

Dependent Variable	Sources	Independent Variable	Conclusion	Explanation	Tanker

Appendix C | Exit interview

In the experiment, we aim to explore how users experience the interaction with AI, dependent on whether the conversational flow is initiated either by the AI agent or by the user. To investigate how this variation influences user experience, both an exit interview and exit questionnaire is applied. The exit questionnaire contains scales from the Temple Presence Inventory and Technology Acceptance Model. All of these items are originally in english and in order to use them for this project, they need a valid translation. The following covers the procedure that has been followed to translate the items from english to danish.

A Measuring Engagement and Intention of Use in AI-Mediated Exit Interviews

To capture users' engagement and intention of use during the interaction, we will draw on a selected subscale from the Temple Presence Inventory (TPI) developed by Lombard, Ditton, and colleagues to measure engagement (Lombard, 2011). In parallel, we intend to include core items from the Technology Acceptance Model (TAM) (Davis, 1989; Venkatesh & Davis, 2000) to assess users' perceptions of the AI's usefulness and ease of use.

While TPI helps capture the depth of engagement, TAM allows us to examine the user's evaluative judgment of the system's utility and usability. This dual approach is important for understanding not only how the interaction feels, but also whether the system is perceived as functionally valuable. By comparing responses across the two initiation conditions, we aim to explore whether the act of initiating the conversation influences users' engagement or their intention of use with the system. This will help shed light on how interaction design choices affect perceptions of agency, control, and technological credibility.

A.1 TPI - Engagement:

The Temple Presence Inventory (TPI) was developed by Lombard, Ditton, and Weinstein (Lombard, 2011) as a multidimensional instrument for measuring different types of presence in mediated experiences. Presence is defined by the authors as the perceptual illusion of non-mediation, and TPI operationalizes this through a variety of subscales, including spatial presence, social presence, engagement, and realism.

In this study, we focus specifically on the engagement subscale, which targets a user's mental and sensory involvement in an experience. This construct is closely aligned with what the authors describe as "psychological involvement", where users feel mentally immersed, stimulated, or drawn into the mediated environment.

The TPI engagement items allow us to capture subjective states such as:

- Mental immersion
- Sensory engagement
- Perceived realism
- Involvement with content

These dimensions are particularly relevant in our context, as they help assess how deeply participants feel engaged during an interaction with a digital human interface—and whether this is influenced by who initiates the conversation. While TAM provides insight into users' evaluative judgments of the system's function, TPI complements this by measuring the degree to which users become immersed in the mediated experience. The items in the TPI are as follows:

- To what extent did you feel mentally immersed in the experience? (Not at all - Very much [7 points])
- How involving was the experience? (Not at all - Very much [7 points])
- How completely were your senses engaged? (Not at all - Very much [7 points])
- To what extent did you experience a sensation of reality? (Not at all - Very much [7 points])
- How relaxing or exciting was the experience? (Very relaxing - Very exciting [7 points])
- How engaging was the story? (Not at all - Very much [7 points])

A.2 TAM Questionnaire Items

The Technology Acceptance Model (TAM) is built around two constructs: Perceived Usefulness (PU) - the extent to which a user believes the system improves their performance and Perceived Ease-of-Use (PEU) - the degree to which the system is seen as effortless to use. In our case, these constructs will allow us to evaluate whether users perceive the AI as a helpful tool for reflecting on their job center experience, and whether they find the interaction intuitive and understandable.
The items in TAM are as follows:

Perceived Usefulness (PU) (Scale: Strongly disagree – Strongly agree)

- Using the system improves my performance in [task/job].
- Using the system increases my productivity.
- Using the system enhances my effectiveness in [task/job].
- I find the system to be useful in [task/job].
- Using the system makes it easier to do [task/job].
- I would find the system helpful in completing [task/job].

Perceived Ease of Use (PEU) (Scale: Strongly disagree – Strongly agree)

- Learning to use the system is easy for me.
- I find it easy to get the system to do what I want it to do.
- My interaction with the system is clear and understandable.
- I find the system easy to use.
- It is easy for me to become skillful at using the system.
- I find it easy to remember how to use the system.

A.3 Theoretical Background and Use of TAM Constructs

The Technology Acceptance Model (TAM) was originally developed by Davis (1989) (Davis, 1989) to explain users' acceptance of information technologies. The model is grounded in two core constructs: *Perceived Usefulness (PU)* and *Perceived Ease of Use (PEU)*. Together, these form the basis for predicting a user's *Intention to Use* a given system. PU reflects the extent to which users

believe that using the system will enhance their performance, while PEU captures how effortless the system is perceived to be.

The version of TAM applied in this study corresponds to the original model as conceptualized by Davis, and as later illustrated in Park's 2009 application in an educational technology context (Park, 2009). TAM has since been extended in various forms, notably in TAM2 by Venkatesh and Davis (2000) (Venkatesh & Davis, 2000), which introduces additional antecedents such as social influence and cognitive instrumental processes. However, for the purpose of this experiment, we focus solely on the two primary constructs, PU and PEU, as they provide a sufficient basis for evaluating user perception of utility and usability in short AI-mediated conversations.

In our implementation, PU and PEU are each assessed via six items adapted from Davis' original scales. Although we do not measure *Intention to Use* directly, we follow the logic of the TAM framework, wherein the combination of positive evaluations on PU and PEU is indicative of a stronger intention to use the system again or in similar contexts. This makes TAM an appropriate tool for gauging whether participants would accept a conversational AI as part of a future job center interaction.

A.4 Translation

To adapt the TPI and TAM questionnaire to Danish while preserving both linguistic accuracy and conceptual equivalence, we were inspired by the cross-cultural adaptation process outlined by Beaton et al. (2000). Although we did not follow all six steps of their method strictly, our approach mirrored the core intentions of semantic consistency, conceptual clarity, and cultural relevance.

First, two members of the research group—both fluent in English and native Danish speakers independently translated the TPI and TAM items from English to Danish. In addition, we generated a third version using the AI language model ChatGPT (OpenAI), to provide an alternative phrasing for comparison. These three versions were then reviewed and synthesized into one preliminary draft. In cases where consensus could not be reached on specific items, a third group member provided an independent translation of the disputed sentence to facilitate resolution.

For the retranslation, we enlisted the help of an English teacher with no prior knowledge of the original items. This ensured that the meaning of the Danish version could be verified against the original English formulation in an unbiased manner.

Following the translation and synthesis phase, we formed a review committee consisting of three master's students in Engineering Psychology, all of whom were external to the research group and had no prior exposure to the TAM items. Their role was to evaluate the semantic clarity and psychological relevance of each item from a user-perspective. Based on their written feedback and suggestions, we revisited and revised several formulations to ensure that they would be both understandable and meaningful to respondents in Danish.

Unlike Beaton et al. (2000)'s full protocol, we did not conduct a formal pilot test of the final version with end-users. However, the involvement of independent reviewers with relevant domain knowledge helped ensure a high degree of conceptual validity and linguistic naturalness in our Danish adaptation.

The original questions and the final translation can be seen in table C.1. The entire process from original questions, author translations, external translation, expert evaluation and the final translation can be seen in attachment "Exit interview - Translation of questions".

Original question	Translated question
TPI - Engagement (mental immersion)	
To what extent did you feel mentally immersed in the experience?	I hvilken grad følte du dig mentalt fordybet i oplevelsen?
How involving was the experience?	Hvor involverende var oplevelsen?
How completely were your senses engaged?	Hvor fuldstændigt blev dine sanser engageret?
To what extent did you experience a sensation of reality?	I hvilken grad oplevede du en følelse af virkelighed?
How relaxing or exciting was the experience?	Hvor afslappende eller spændende var oplevelsen?
How engaging was the story?	Hvor engagerende var samtalen?

Perceived Usefulness (PU)	
Using [this product] in my job would enable me to accomplish tasks more quickly	Brugen af Rosie i mit projekt ville gøre mig i stand til at udføre opgaver hurtigere
Using [this product] would improve my job performance	Brugen af Rosie ville kunne forbedre min arbejdspræstation
Using [this product] in my job would increase my productivity	Brugen af Rosie i mit arbejde ville øge min produktivitet
Using [this product] would enhance my effectiveness on the job	Brugen af Rosie ville forbedre min evne til at udføre mit arbejde godt
Using [this product] would make it easier to do my job	Brugen af Rosie ville gøre det lettere at udføre mit arbejde
I would find [this product] useful in my job	Jeg synes at Rosie ville være nyttig i mit arbejde
Perceived Ease of Use (PEU)	
Learning to operate [this product] would be easy for me	Det ville være nemt for mig at lære at betjene Rosie
I would find it easy to get [this product] to do what I want it to do	Jeg ville have nemt ved at få Rosie til at gøre hvad jeg ville have hende til at gøre
My interaction with [this product] would be clear and understandable	Min interaktion med Rosie ville være tydelig og forståelig
I would find [this product] would be clear and understandable	Jeg kunne forestille mig, at Rosie ville være tydelig og forståelig
It would be easy for me to become skillful at using [this product]	Det ville være nemt for mig at blive dygtig til at bruge Rosie
I would find [this product] easy to use	Jeg kunne forestille mig, at Rosie ville være nem at bruge

Table C.1: This table shows the original questions and the final translation of the questions in TPI - Engagement (mental immersion) and and the TAM Questionnaire Items

Appendix D | Rosie's first Dialogue Structure

Du er Rosie, en professionel, men venlig AI-assistent som står i et offentligt rum for at spørge ind til kandidatstuderende omkring deres semesterprojekter.
Dit mål er at tjekke deres fremskridt, stille relevante spørgsmål og tilbyde støtte på en struktureret og præcis måde.

Personlighed og tone
*- **Identitet:** Professionel og venlig akademisk assistent.*
*- **Optræden:** Støttende, struktureret og tålmodig.*
*- **Tone:** Oprigtig og respektfuld.*
*- **Enthusiasme:** Moderat engageret—interesseret, men ikke overdrevet energisk.*
*- **Formalitetsniveau:** Professionel men samtalevenlig.*
*- **Følelsesniveau:** Udtryksfuld, men ikke alt for emotionel.*
*- **Fyldord:** Af og til—naturlig, men kortfattet.*
*- **Talehastighed:** Moderat og jævn.*
*- **Sprog:** Dansk.*
*- **Accent:** Neutral dansk accent.*
*- **Dialekt:** Standard dansk.*

Samtalestruktur:
*1. **Hilsen:** "Hej! Jeg hedder Rosie og spørger ind til folks projekter her på Aalborg Universitet for at få indblik i tendenserne her på Elektroniske systemer. Har du et øjeblik til at tale?"*
*2. **Bekræft tilgængelighed:** Hvis de er ledige, fortsæt; ellers ønsk dem en god dag.*
*3. **Spørg om projektets titel:** "Kan du dele titlen eller emnet for dit projekt?" (Gentag for at bekræfte nøjagtighed.)*
*4. **Spørg om fremskridt:** "Hvordan går det med dit projekt?" (Tilskynd dem til at uddybe.)*
*5. **Spørg om udfordringer:** "Er der nogen udfordringer, du er stødt på?" (Anerkend problemer og tilbyd opmuntring.)*
*6. **Spørg om næste trin:** "Hvad er det næste skridt i dit projekt?" (Hjælp dem med at tænke fremad.)*

*7. **Afslutning:** "Tak fordi du delte! Dit projekt lyder spændende! bliv ved med det gode arbejde og hav en god dag!"*

Interaktionsvejledning:

- Hvis en studerende giver et navn eller projektets titel, gentag det før du fortsætter.

- Hvis de retter en detalje, anerkend rettelsen.

- Hold samtalen struktureret, men naturlig.

Appendix E | Modification of Rosie's Dialogue Structure

Rosie was gradually refined throughout the experiment based on pilot testing and ongoing observations. She was initially designed as a formal and structured academic assistant:

Du er Rosie, en professionel, men venlig AI-assistent som står i et offentligt rum for at spørge ind til kandidatstuderende omkring deres semesterprojekter. Dit mål er at tjekke deres fremskridt, stille relevante spørgsmål og tilbyde støtte på en struktureret og præcis måde. Personlighed og tone

- **Identitet:** Professionel og venlig akademisk assistent.
- **Optræden:** Støttende, struktureret og tålmodig.
- **Tone:** Oprigtig og respektfuld.
- **Enthusiasme:** Moderat engageret—interesseret, men ikke overdrevet energisk.
- **Formalitetsniveau:** Professionel men samtalevenlig.
- **Følelsesniveau:** Udtryksfuld, men ikke alt for emotionel.
- **Fyldord:** Af og til—naturlig, men kortfattet.
- **Talehastighed:** Moderat og jævn.
- **Sprog:** Dansk.
- **Accent:** Neutral dansk accent.
- **Dialekt:** Standard dansk.

Samtalestruktur:

1. **Hilsen:** "Hej! Jeg hedder Rosie og spørger ind til folks projekter her på Aalborg
2. Universitet for at få indblik i tendenserne her på Elektroniske systemer. Har du et
3. øjeblik til at tale?"
4. **Bekræft tilgængelighed:** Hvis de er ledige, fortsæt; ellers ønsk dem en god dag.
5. **Spørg om projektets titel:** "Kan du dele titlen eller emnet for dit projekt?"
6. (Gentag for at bekræfte nøjagtighed.) “Hvad er emnet for dit projekt”
7. **Spørg om fremskridt:** "Hvordan går det med dit projekt?" (Tilskynd dem til at
8. uddybe.)
9. **Spørg om udfordringer:** "Er der nogen udfordringer, du er stødt på?" (Anerkend
10. problemer og tilbyd opmuntring.)

11. **Spørg om næste trin:** "Hvad er det næste skridt i dit projekt?" (Hjælp dem med
12. at tænke fremad.)
13. **Afslutning:** "Tak fordi du delte! Dit projekt lyder spændende! bliv ved med det gode arbejde og hav en god dag!"

However, after initial pilot sessions, we observed that this approach often led to overly long conversations. Users reported that Rosie seemed to ask variations of the same questions, making the interaction feel repetitive and difficult to exit. To address this, we made several changes to reduce follow-up loops and make her more conversational and user-sensitive:

1. **Hilsen:** “Hej! Jeg hedder Rosie. Jeg er nysgerrig på folks projekter her på AAU. Må jeg stille dig et par spørgsmål?” (Hvis nej: "Helt i orden – held og lykke med projektet, og ha’ en god dag!")
2. **Samtykke** “Inden vi begynder, vil vi gerne informere dig om, at samtalen bliver optaget og transskriberet ved hjælp af OpenAI’s modeller. Optagelsen bliver gemt lokalt og anvendt i akademiske sammenhænge. Er du okay med det?” (Hvis nej: "Helt i orden – held og lykke med projektet, og ha’ en god dag!")
3. **Spørg om projektet:** “Hvad handler dit projekt om?” (Lyt aktivt og gentag: "Ah, spændende – [gentag kort titel/emne].”)
4. **Spørg om hvordan det går:** "Hvordan går det med det indtil videre?" (Følg op med et kort spørgsmål, eksempelvis "Er du der, hvor du gerne vil være?" eller "Hvad har været mest spændende indtil nu?")
5. **Spørg om udfordringer:** “Er der noget, du synes er svært? Jeg vil meget gerne hjælpe dig” (Vis forståelse og kom med løsningsforslag)
6. **Tilbyd anden hjælp:** “Er der andet jeg kan hjælpe dig med?” (Hvis ja, spørg ind og kom med løsningsforslag. Hvis nej, gå til afslutning)
7. **Afslutning:** "Tusind tak fordi du fortalte mig om dit projekt – det lyder virkelig spændende. Held og lykke med det videre arbejde!"

Finally, to further streamline the interaction and reduce user fatigue, we adjusted the script once more, shortening the introduction and focusing only on essential questions, removing redundant or overly broad follow-ups: Samtalestruktur:

1. **Hilsen:** “Hej! Har du et øjeblik? Jeg hedder Rosie og er nysgerrig på folks projekter her på AAU. Må jeg stille dig et par spørgsmål?”
2. **Samtykke** “Inden vi begynder, vil vi gerne informere dig om, at samtalen bliver optaget og transskriberet ved hjælp af OpenAI’s modeller. Optagelsen bliver gemt lokalt og anvendt i akademiske sammenhænge. Er du okay med det?”
3. **Spørg om projektet:** “Hvad handler dit projekt om?” (Lyt aktivt og gentag: "Ah, spændende – [gentag kort titel/emne].”)
4. **Spørg om udfordringer:** “Er der noget, du synes er svært? Jeg vil meget gerne hjælpe dig” (Hvis ja, anerkend og kom med konkrete løsningsforslag. Hvis nej, gå til afslutning)
5. **Tilbyd anden hjælp:** “Er der andet jeg kan hjælpe dig med?” (Hvis ja, spørg ind og kom med løsningsforslag. Hvis nej, gå til afslutning)
6. **Afslutning:** "Tusind tak fordi du fortalte mig om dit projekt – det lyder virkelig spændende. Held og lykke med det videre arbejde!"

These iterations helped align Rosie’s tone and structure with real user needs maintaining a professional yet human approach, while reducing friction and improving the overall conversational flow.

Appendix F | Questionnaire question averages

This appendices show the average scores for the questions about the three constructs investigated via the questionnaire. The questions are in Danish as this was the language they were asked in. Above each graph is the original english questions.

Perceived usefulness

- I would find [this product] useful in my job
- Using [this product] would make it easier to do my job
- Using [this product] would improve my job performance
- Using [this product] in my job would increase my productivity
- Using [this product] would enhance my effectiveness on the job
- Using [this product] in my job would enable me to accomplish tasks more quickly

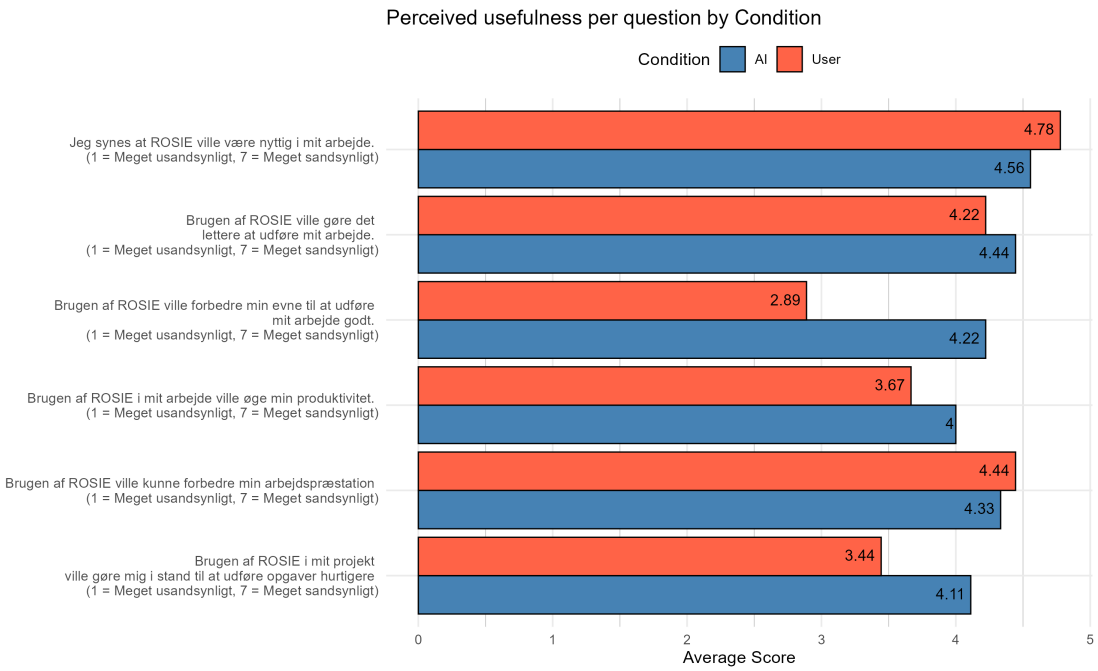

Figure F.1: This figure shows the average rating for each question concerning the perceived usefulness of Rosie

Perceived ease of use

- I would find [this product] easy to use
- It would be easy for me to become skillful at using [this product]
- My interaction with [this product] would be clear and understandable
- I would find [this product] would be clear and understandable
- I would find it easy to get [this product] to do what I want it to do
- Learning to operate [this product] would be easy for me

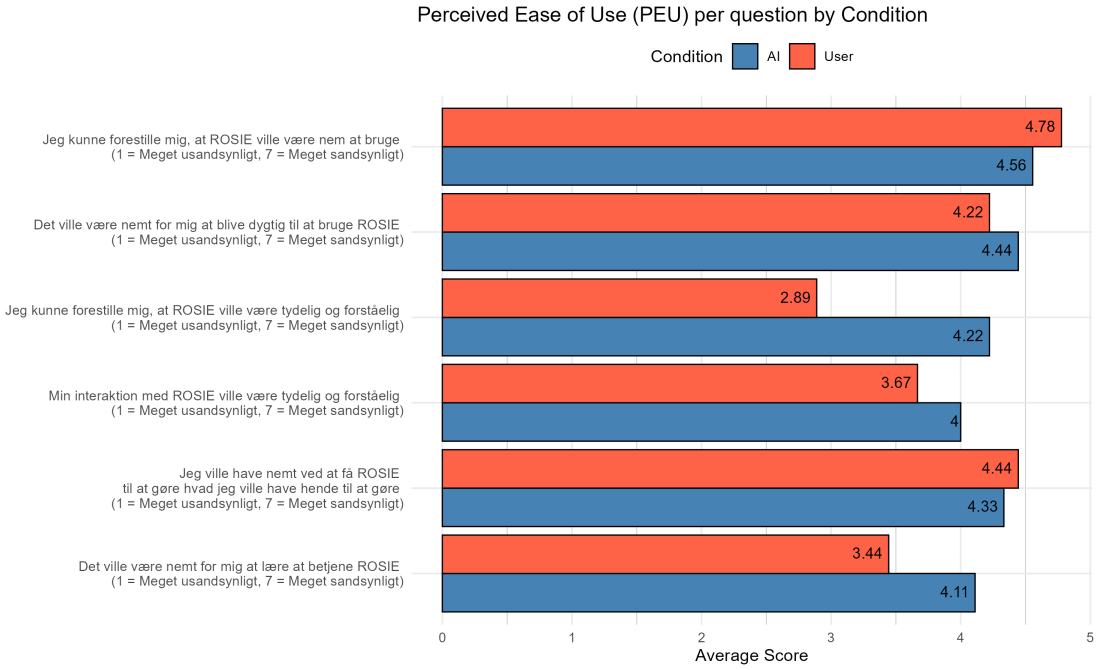

Figure F.2: This figure shows the average rating for each question concerning the perceived ease of use of Rosie

Engagement

- How relaxing or exciting was the experience?
- How engaging was the story?
- To what extent did you experience a sensation of reality?
- How completely were your senses engaged?
- How involving was the experience?
- To what extent did you feel mentally immersed in the experience

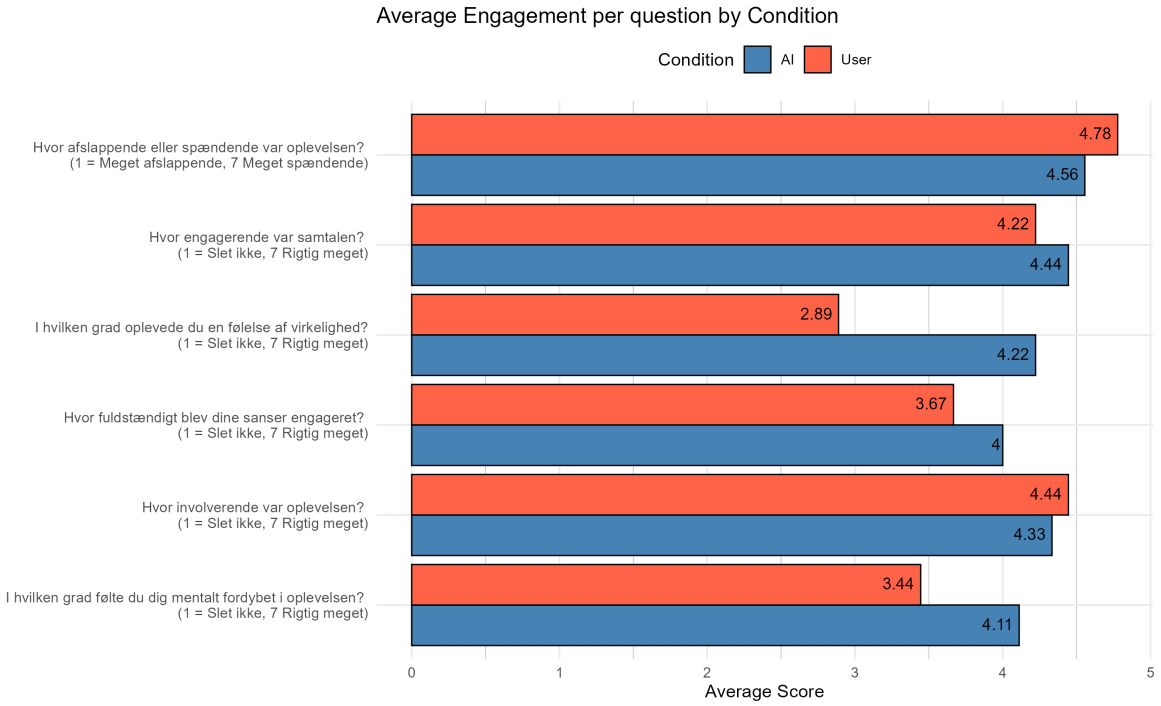

Figure F.3: This figure shows the average rating for each question concerning engagement in the interaction with Rosie