Opleiding Informatica & Economie

More Human, Less Bot: How Social Presence Impacts Satisfaction, Trust, and Reuse Intention in Customer Support Chatbots

Christiaan van Driel

Supervisors:
Natalia Amat Lefort & Yingjie Fan

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

24/07/2024

Abstract

This study investigates the influence of chatbot design cues on user perceptions of Social Presence and how Social Presence dimensions impact user experience in customer support interactions. A randomized controlled trial online experiment was conducted where participants viewed one of three pre-recorded customer support scenarios featuring identical conversations but varying chatbot designs (no avatar, static avatar, dynamic avatar with facial expressions). Interestingly, the experiment found no significant differences in user perceptions of human-likeness, naturalness, lifelikeness, or professionalism across the three chatbot designs. However, the study revealed positive relationships between user ratings of human-likeness, naturalness, and professionalism with user satisfaction and trust in the chatbot. These findings suggest that while users may not consciously distinguish between different levels of anthropomorphism, Social Presence dimensions do influence user experience with chatbots. This highlights the importance of considering user experience in chatbot design, with a focus on user satisfaction, trust, and emotional connection, even if users do not explicitly perceive subtle differences in terms of anthropomorphic features.

Content

1. Introduction

Customer service is undergoing a digital revolution, with chatbots – computer programs designed to simulate conversation with human users – rapidly becoming widespread (Adamopoulos & Moussiades, 2020). While chatbots offer benefits such as 24/7 availability and faster response times (Adam et al., 2020a), limitations in handling complex requests and expressing empathy can lead to frustration (Luger & Sellen, 2016). Gartner predicts that by 2027, a quarter of organizations will rely on chatbots as their primary customer service channel (Gartner, 2022). This highlights the critical need to design chatbots that deliver positive user experiences.

One key factor influencing user experience may be a chatbot's Social Presence, the feeling of interacting with a real person. Recent research suggests that the visual design of chatbots can influence consumer perception and satisfaction (Klein & Martinez, 2022). However, a critical gap remains: how does a chatbot's Social Presence, influenced by anthropomorphic design cues (i.e., features resembling a human), impact consumer experience in customer support interactions?

This study investigates the influence of a chatbot's visual design on user perceptions of Social Presence and how Social Presence in turn impacts user experience. Subsequently, this research explores how different Social Presence dimensions (e.g., human-likeness, naturalness) impact Consumer Satisfaction, Trust, and reuse intention in the context of customer support chatbots.

We conducted a randomized controlled trial where participants were exposed to one of three customer support chatbots featuring identical conversations but varying levels of anthropomorphism (no avatar, static avatar, dynamic avatar with facial expressions). Drawing upon Social Presence Theory (Short et al., 1976), we hypothesized that chatbots with anthropomorphic design cues (e.g., human-like avatars and dynamic facial expressions) would create a greater sense of Social Presence, leading consumers to perceive the interaction as more natural and engaging.

Interestingly, the experiment found no significant differences in user perceptions of human-likeness, naturalness, lifelikeness, or professionalism across the three chatbot designs. Despite this unexpected result, the study revealed positive relationships between user ratings of these Social Presence dimensions and user satisfaction and trust in the chatbot. These findings suggest that while users may not consciously distinguish between different levels of anthropomorphism, Social Presence dimensions do influence user experience with chatbots. This highlights the importance of considering user experience in chatbot design, with a focus on user satisfaction and trust, even if users do not explicitly perceive subtle differences in anthropomorphic features.

The results offer valuable insights that bridge the gap between theory and real-world application. From a theoretical standpoint, this study extends the Social Presence Theory by testing its applicability in the context of human-computer interaction with customer support chatbots. By establishing the link between anthropomorphic design cues, Social Presence, and consumer outcomes, this research sheds light on how chatbots can be designed to foster more positive consumer experiences.

In practical terms, the findings can guide the development and implementation of customer support chatbots that leverage Social Presence to enhance consumer experience. Understanding the design elements that influence consumer perceptions of Social Presence allows companies to optimize their chatbots for a more natural and engaging customer service experience. This can lead to improved consumer satisfaction and a more positive brand image.

The structure of the paper is as follows. First, we explain the theoretical framework and context of this research in the literature review section. Subsequently, we develop our hypotheses and present the study's proposed research model in section three. Section four describes the research process and methodology applied. Section five details the empirical results of the analysis. Finally, the last section discusses the results of the research, providing implications for theory and practice, limitations and directions for future research, and conclusions.

2. Literature review

This study investigates the influence of a chatbot's visual design on user perceptions of Social Presence and how Social Presence in turn impacts user experience. Araujo (2018) conducted a similar study that focused on exploring the effects of interacting with disembodied agents. These chatbots did not have profile pictures and interacted with users solely through text (Araujo, 2018). The study investigated a range of customer outcomes, including attitudes, satisfaction, and emotional connection towards the company (Araujo, 2018).

In another related study, Konya-Baumbach et al. (2023) assessed the effectiveness of chatbot anthropomorphism during interactions with users. This study implemented human-like linguistic cues in chatbots, generating varying levels of anthropomorphism (Konya-Baumback et al., 2023). Konya-Baumbach et al. (2023) found that anthropomorphism had positive effects on satisfaction, trust, and other factors such as purchase intention and word-of-mouth in the shopping experience. Social Presence was identified as the mechanism driving these effects, underscoring its importance in customer-chatbot interactions (Konya-Baumback et al., 2023).

In comparison to the studies done by Araujo (2018) and Konya-Baumbach et al. (2023), this study investigates the visual design and appearance of chatbots instead of focusing on the anthropomorphism design used in text interactions. By doing so, this research adds a new dimension to our understanding of how visual elements contribute to Social Presence and user experience in chatbot interactions.

2.1. Use of Chatbots for Customer Support

The evolution of information and communication technologies has had a profound impact on personal lifestyles as well as business operations. This has resulted in businesses needing to enhance their services and offer accessible and convenient communication channels for their users (Kwangsawad & Jattamart, 2022). While the use of digital self-service options within customer support, such as customer webpages and smartphone apps, is rising there is still a demand for skilled customer support agents (Følstad et al., 2014). Chatbots for customer support may resemble users' conversations with real-life customer support agents. This resemblance leads to chatbots being perceived as more accessible by users than web page interactions (Følstad & Skjuve, 2019).

Interacting with users through live channels, including chatbots, has been gaining popularity in the digital customer support environment. These channels are used for several reasons such as requesting information about certain products or addressing technical problems the customers might be experiencing (Adam et al., 2020). Although chat services and the use of chatbots have become more popular there are still problems when it comes to meeting the customer's expectations. Businesses benefit from cost and time savings when they use chatbots. However, in practice, chatbots usually fail short of users' expectations, which makes them less likely to comply with the chatbot's instructions (Adam et al., 2020). It is therefore important to bridge the gap between chatbots and real-life customer support employees to meet the customer's expectations.

2.2. Social Presence Theory

One way to bridge this gap is by analyzing a chatbot's Social Presence. Social Presence refers to "the feeling that other actors are jointly involved in communicative interaction." (Walther 1992, p. 54). Social Presence theory suggests that the perceived presence of other actors within mediated communication environments influences the quality of said interactions. This theory advocates that communication technologies have various levels of capacity to transmit social presence to individuals who use them. (Short et al., 1976; Walther, 1992).

Short et al. (1976, p. 65) state that electronic media differ in their "capacity to transmit information about facial expression, direction of looking, posture, dress and nonverbal, vocal cues." This can affect the degree of social presence that users experience. In addition, Walther (1992, p. 54-55) notes that "computer-mediated communication, with its paucity of nonverbal elements and backchanneling cues, is said to be extremely low in social presence in comparison to face-to-face communication." It is therefore no surprise that companies aim to increase their chatbot's social presence to make them come across as more realistic and natural (Araujo, 2018).

2. 3 Assessment of Social Presence

Social Presence is a critical component of effective communication, influencing how users perceive and interact with technology. To evaluate the Social Presence of a chatbot in this study, we focus on four key dimensions: Human-likeness, Naturalness, Lifelikeness, and Professionalism.

Human Likeness

Human-likeness refers to the extent to which a chatbot is designed to appear and behave like a human. Smestad (2018, p. 9) defines humanness as "the extent to which an agent is designed to act and appear human [...] encompassing the objectively established human capabilities (having eyes, a face or the ability to respond politely)" (Meyer et al. 2016). This definition underscores the importance of incorporating human-like attributes, such as facial features and polite language, to enhance user experience and perception.

Beyond physical appearance, language characteristics also contribute significantly to human-likeness. Industry reports indicate a strong preference among users for chatbots with human-like qualities, including friendliness (Drift, 2018). Haugeland et al. (2022) further emphasize the role of human-likeness in reinforcing perceptions of anthropomorphism and social presence. Anthropomorphism, as described by Araujo (2018) and Nass & Moon (2000), refers to the attribution of human-like qualities to non-human entities. Thus, human-likeness, encompassing physical attributes and language characteristics, is crucial in establishing social presence within chatbot interactions.

Naturalness, Lifelikeness, and Professionalism

To complement the assessment of human-likeness, we also consider three additional dimensions: naturalness, lifelikeness, and professionalism. These dimensions collectively contribute to the overall perception of a chatbot as a genuine and competent conversational partner.

- ***Naturalness***: To create a sense of real-world interaction, chatbots should exhibit natural conversational flow. Atiyah (2019) highlights the importance of assigning unique names and personalities to chatbots to foster a feeling of personal connection. This approach aligns with the Cambridge dictionary's definition of naturalness as "the quality of being real and not influenced by other people" (2024).
- ***Lifelikeness:*** This dimension extends beyond naturalness to encompass the overall impression of a chatbot as a living, responsive entity. Similar to naturalness, lifelikeness is closely tied to human-likeness, emphasizing the importance of designing chatbots that evoke a sense of realism. The Cambridge Dictionary defines lifelikeness as "used to describe something that appears real or very similar to what is real" (2024).
- ***Professionalism*** In customer service interactions, users expect chatbots to exhibit professionalism, characterized by competence, skill, organization, and a serious demeanor (Goodwin & Smith, 1990). The Cambridge Dictionary defines professionalism as "having the qualities that you connect with trained and skilled people" (2024).

By considering these four dimensions, we aim to provide a comprehensive assessment of Social Presence in chatbots and its influence on user experience.

2.4. Importance of Chatbot Design

To bridge this gap between chatbots and real-life customer support employees chatbot design is of significant importance. Chatbot design significantly influences consumer satisfaction, behavior, and perception of the chatbot (Jain et al., 2018). Research has shown that positively influencing consumer satisfaction, behavior, and perception through chatbot design can result in better customer engagement with the chatbot (Sheehan et al., 2020).

When designing a chatbot, and more importantly a chatbot identity, it is necessary to think about the chatbot's attributes which can have impact on users. One of these attributes is the chatbot’s assigned gender. According to a study by Toader et al. (2019), female chatbots with a high level of anthropomorphism and social interaction significantly shape positive responses from customers, even when errors occur. In addition, female chatbots were more frequently forgiven for making mistakes compared to their male counterparts. Given these pronounced differences in customer response based on chatbot gender, Toader et al. (2019, p. 19) “advise practitioners to strongly consider the deployment of female virtual assistants for customer case interactions in a retail context.”

2.5. Customer Outcomes

The most important aspect of a chatbot is to give customers satisfactory help with their questions and problems. How a chatbot can handle these questions and problems, is crucial to the business. If a consumer has a positive experience, they may be inclined to engage further with the business’s products or services (Otto et al., 2019). This in turn will lead to higher profits (Mittal et al., 2017). Chatbots can also save time and personnel costs by either taking over part of the human work or being supportive to the human customer care workers (Khwaja & Yang, 2022).

If a consumer has a negative experience, leading to a negative outcome, this will have the opposite effect (Mittal et al., 2017; Otto et al., 2019). A chatbot which can create satisfactory consumer outcomes and negates unsatisfactory consumer outcomes can therefore add real value to a business. Some of the key metrics to determine if a chatbot functions well are Consumer Satisfaction, Trust, and the consumer’s intention to use the chatbot again. These metrics can be defined as the following:

Customer satisfaction: According to Anderson and Sullivan (1993), customer satisfaction is a general assessment of how well a product or service performs in relation to the expectations of the customer. Other literature gives a similar definition, they define customer satisfaction as the assessment of a good or service by the customer following purchase, use, and consumption (Silva et al., 2023; Türkyilmaz & Özkan, 2007; Kim, 2019).

The foundation of ***Trust*** is the confidence that one's weaknesses will not be used against them in a potentially dangerous online scenario (Silva et al., 2023; Aljazzal et al., 2010; Corritore et al., 2003). *Trust* is essential because it affects a customer's propensity to accept the information to accept the information the chatbot presents, follow their recommendations, and take advantage of the benefits that come with using it (Toader et al., 2019; Hancock et al., 2011)

Reuse intention can be defined as the likelihood of customers continuing to use a product or service based on their previous experiences (Silva et al., 2023). In the context of this study, this means that Reuse intention is the likelihood of customers using the chatbot.

Having defined the consumer outcomes (namely customer satisfaction, trust, and reuse intention) and articulated the framework for Social Presence in the context of chatbots, which includes dimensions such as human likeness, naturalness, lifelikeness, and professionalism, we now transition to the third section. This section formulates the hypotheses and introduce the research model using the framework described in this section.

3. Hypotheses development and Research model

This section outlines the key hypotheses guiding this research project. Subsection 3.1. introduces the hypotheses related to the first research objective of this study: to explore the influence of anthropomorphic design cues on perceived Social Presence in customer support chatbots. The remaining subsections present the hypotheses linked to the second research goal of this study: to examine the impact of each Social Presence dimension on Consumer Satisfaction, Trust, and Reuse intention.

3.1. Anthropomorphic Design Cues and Social Presence

Social Presence Theory (SPT) (Short et al., 1976) notes that consumers perceive interactions with entities exhibiting human-like qualities as more natural and engaging, leading to a stronger sense of Social Presence. In the context of chatbots, this translates to consumers feeling as if they are interacting with another person rather than a machine. Drawing upon SPT (Short et al., 1976), we suggest that chatbots with anthropomorphic design cues (e.g., human-like avatars) can trigger these associations. As a result, we expect chatbots with such design elements to be perceived as having a higher degree of Social Presence compared to those without.

3.1.1. No avatar vs. (Static) Human-like avatar

The presence of a human-like avatar in a chatbot can function as a powerful visual cue, suggesting to the consumer that they are interacting with another person rather than a machine. This, in turn, can increase the perceived human likeness of the interaction. Furthermore, a well-designed avatar can be crafted to appear natural and lifelike, further enhancing the overall sense of Social Presence. The avatar can also project professionalism, leading users to perceive the chatbot as a more trustworthy and reliable source of information. In essence, the human-like qualities embedded within the avatar can contribute to a more positive consumer experience on multiple levels. Therefore, we propose the following hypotheses:

H1: Perceived Social Presence will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.

We can break down H1 into more specific hypotheses for each Social Presence dimension.

- H1a: Perceived Human-likeness will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.

- H1b: Perceived Naturalness will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.
- H1c: Perceived Lifelikeness will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.
- H1d: Perceived Professionalism will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.

3.1.2. No avatar vs. Human-like avatar with dynamic facial expressions

Facial expressions are a fundamental aspect of human communication, conveying emotions, intentions, and even personality traits without a single word spoken (Hess, 2020). When interacting with others, humans tend to pay close attention to nonverbal cues, including facial expressions (Hess, 2020). Research shows that recognizing emotions in facial expressions is not only automatic but also plays a crucial role in social interactions, fostering empathy and understanding (Hess, 2020).

Media Richness Theory (MRT) (Daft & Lengel, 1984) suggests that communication channels with richer nonverbal cues enhance user understanding and satisfaction. Facial expressions are a key form of nonverbal communication. Aligning with Media Richness Theory, we propose that chatbot avatars with dynamic facial expressions will be perceived as richer and more natural compared to those with static expressions or those without an avatar. By providing nonverbal cues that mirror human interaction, this increased richness can contribute to a stronger sense of Social Presence. This leads us to the hypotheses H2 and H3 (section 3.1.3).

H2: Perceived Social Presence will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.

We can break down H2 into more specific hypotheses for each Social Presence dimension.

- H2a: Perceived Human-likeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.
- H2b: Perceived Naturalness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.
- H2c: Perceived Lifelikeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.

- H2d: Perceived Professionalism will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.

3.1.3. Static Human-like avatar vs. Human-like avatar with dynamic facial expressions

H3: Perceived Social Presence will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.

We can break down H3 into more specific hypotheses for each Social Presence dimension.

- H3a: Perceived Human-likeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.
- H3b: Perceived Naturalness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.
- H3c: Perceived Lifelikeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.
- H3d: Perceived Professionalism will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.

3.2 Social Presence and Consumer Satisfaction

In the context of chatbots, Social Presence can be created through design elements that evoke human-like qualities. The Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1986) provides a valuable framework for understanding how Social Presence might influence consumer experience. ELM suggests that individuals process information through two primary routes: the central route and the peripheral route. The central route involves thoughtful consideration of the information itself, while the peripheral route relies on simpler cues to form judgments.

Customer support chatbot interactions often involve the peripheral route, as consumers may not need to engage in deep analysis of the information exchanged. In these situations, peripheral cues like a chatbot's appearance can play a significant role in shaping consumer perception. Building on ELM, we suggest that a chatbot with a high degree of Social Presence, achieved through anthropomorphic design cues, can trigger positive associations with human interaction.

These positive associations, even without extensive information processing, can lead to a more positive consumer experience. Thus, we propose the following hypotheses:

H4: Perceived Social Presence in a customer support chatbot will be positively associated with Consumer Satisfaction.

We can break down H4 into more specific hypotheses for each Social Presence dimension.

- H4a: Perceived Human-likeness in a customer support chatbot will be positively associated with Consumer Satisfaction.
- H4b: Perceived Naturalness in a customer support chatbot will be positively associated with Consumer Satisfaction.
- H4c: Perceived Lifelikeness in a customer support chatbot will be positively associated with Consumer Satisfaction.
- H4d: Perceived Professionalism in a customer support chatbot will be positively associated with Consumer Satisfaction.

3.3 Social Presence and Trust

Building trust with consumers is essential for successful technology adoption (Lukyanenko et al., 2022), and Social Presence can play a crucial role in this process. When interacting with a machine, consumers may experience feelings of uncertainty or apprehension (Dekkal et al., 2023). Social Presence can foster a more trusting relationship by creating a sense of familiarity and reducing this uncertainty. This is supported by Social Identity Theory (SIT) (Tajfel & Turner, 1979), which suggests that individuals are more likely to trust those they perceive as similar to themselves. By exhibiting human-like characteristics, a chatbot can trigger a sense of in-group membership, increasing user trust. Furthermore, interpersonal trust often relies on nonverbal cues that signal sincerity and competence. Leveraging anthropomorphic design cues, chatbots can increase perceived Social Presence and project trustworthiness and reliability. A chatbot that exhibits human-like qualities through design cues may be perceived as more relatable and trustworthy compared to one lacking such features. Therefore, we propose the following hypotheses:

H5: Perceived Social Presence in a customer support chatbot will be positively associated with Trust.

We can break down H5 into more specific hypotheses for each Social Presence dimension.

- H5a: Perceived Human-likeness in a customer support chatbot will be positively associated with Trust.
- H5b: Perceived Naturalness in a customer support chatbot will be positively associated with Trust.
- H5c: Perceived Lifelikeness in a customer support chatbot will be positively associated with Trust.
- H5d: Perceived Professionalism in a customer support chatbot will be positively associated with Trust.

3.4 Social Presence and Reuse Intention

When a customer has a satisfying and productive interaction with a chatbot, they are more likely to choose the same chatbot again for future needs. Social Presence can play a significant role in shaping these positive experiences. A chatbot that exhibits a high degree of Social Presence can foster a more engaging and user-friendly interaction. This is because its human-like characteristics can make it easier to interact with and understand. Consumers may develop a sense of trust and familiarity with the chatbot, making them more likely to choose it over other options for future customer service needs. Therefore, we propose the following hypotheses:

H6: Perceived Social Presence in a customer support chatbot will be positively associated with Reuse Intention.

We can break down H6 into more specific hypotheses for each Social Presence dimension.

- H6a: Perceived Human-likeness in a customer support chatbot will be positively associated with Reuse Intention.
- H6b: Perceived Naturalness in a customer support chatbot will be positively associated with Reuse Intention.
- H6c: Perceived Lifelikeness in a customer support chatbot will be positively associated with Reuse Intention.
- H6d: Perceived Professionalism in a customer support chatbot will be positively associated with Reuse Intention.

3.5 Research Model

The proposed research model visually depicts the relationships between perceived Social Presence dimensions and consumer outcomes (satisfaction, trust, reuse intention). The Social Presence dimensions are the independent variables that influence consumer outcomes. These variables include Human-likeness, Naturalness, Lifelikeness, Professionalism, and Appearance, where Appearance has the moderator variable 'Importance of Appearance.' The model also includes several control variables: Gender, Age, ICT familiarity, and Personality (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to new experiences).

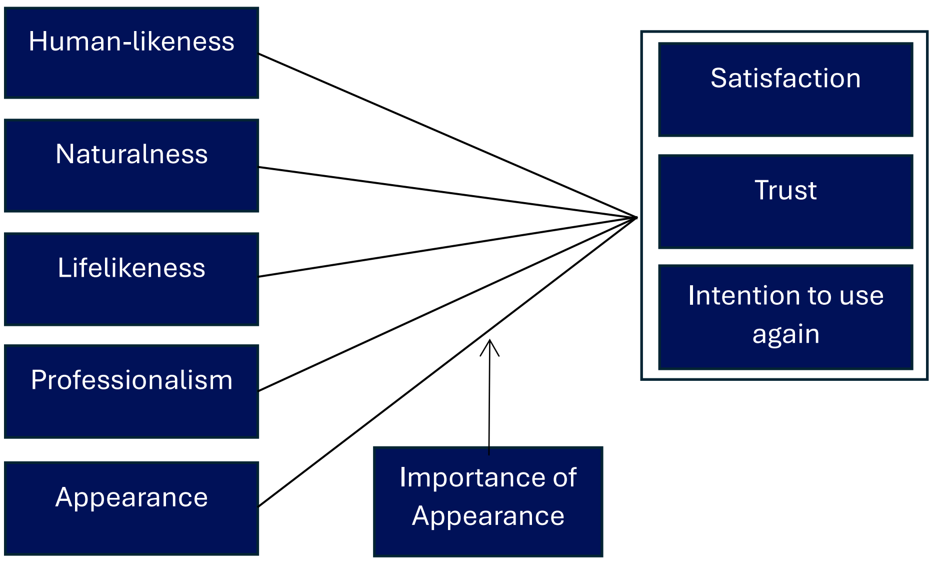

Figure 1. The proposed research model depicts the relationship between perceived social dimensions and consumer outcomes.

4. Methodology

4.1. Experimental Design

This study investigated consumer responses towards the use of chatbots as customer support agents. To achieve this, a between-subjects experiment was conducted. Participants were randomly assigned to watch one of three variations of a pre-recorded customer support scenario. The scenario depicted a customer contacting a chatbot named "Eva" about damaged sunglasses received from an online store. Three animations were created, each featuring an identical conversation but with various levels of anthropomorphism:

- Faceless Avatar: Conversation featuring a faceless avatar.
- Human-like Avatar: Conversation featuring a static, human-like avatar image.
- Human-like Avatar with Dynamic Facial Expressions: Conversation featuring a human-like avatar that displays a range of emotions (e.g., smiling for positive messages, frowning for negative messages) to reflect the conversation's tone.

The decision to use a female persona ("Eva") for the chatbot was informed by previous research suggesting that users tend to rate female virtual agents more favorably (Toad et al., 2019). Using pre-recorded animations ensured minimal variability in the chatbot experience across participants, allowing for reliable comparisons of consumer responses. After watching the animation, participants were asked to rate various aspects of the chatbot.

4.1.2. Chatbot Design

Three chatbot animations were created to explore the effects of anthropomorphic design cues on consumer responses. All animations presented the same conversation between a customer and a customer support chatbot. The key difference between the animations was the level of anthropomorphism embodied by the virtual assistant "Eva."

The first variation of the chatbot presented the conversation featuring a faceless avatar representing the chatbot "Eva" (see Figure 2). This design served as the baseline condition with the lowest level of anthropomorphism.

Figure 2. Faceless avatar, used for the chatbot with the lowest level of anthropomorphism.

The second variation of the chatbot introduced a static human-like avatar image for "Eva" throughout the conversation (see Figure 3). A stock image was used for this purpose (iStock, 2018). This design increased the level of anthropomorphism compared to the baseline condition.

Figure 3. Static human-like avatar.

The third animation featured a human-like avatar with dynamic facial expressions that responded to the emotional tone of the conversation (see Figure 4).

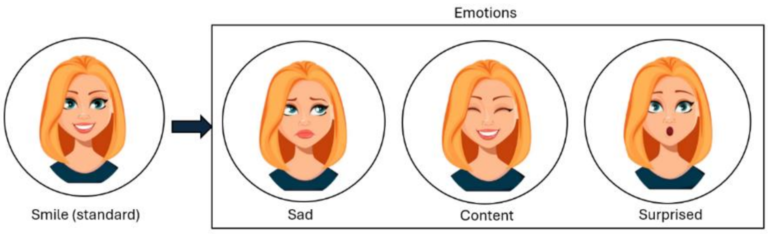

Figure 4. Avatar displaying dynamic facial expressions.

For instance, the avatar would appear sad when responding to negative messages (see Figure 5). When the user in the video responds with a positive text, such as expressing satisfaction with the chatbot's assistance, Eva will display a smile in content whilst responding. This positive animation will be shown briefly before reverting to the default expression. Conversely, if the user responds with a message indicating disappointment or dissatisfaction, Eva will exhibit a sad expression and convey an apology to the user. This design offered the highest level of anthropomorphism among the three variations.

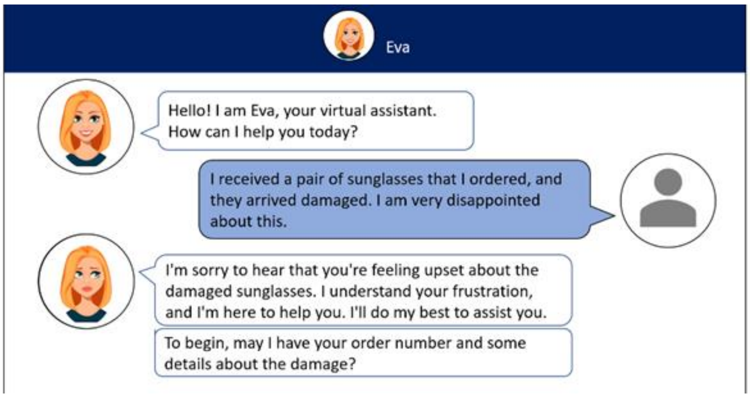

Figure 5. Screenshot of the conversation in which the avatar's facial expressions match the emotions expressed in the textual messages.

By manipulating the level of anthropomorphism across these animations, the experiment aimed to investigate how users perceive and respond to chatbots with varying degrees of human-like characteristics.

4.2. Questionnaire development

To assess participant responses to the chatbot animations, a survey instrument was developed and distributed via multiple online platforms. The survey targeted adults and consisted of 21 total questions of which 17 were used in the model. Table 1 highlights the 17 survey questions which were used for the model and their corresponding model variables. The full list of survey questions can be found in Appendix A.

The survey randomly assigned one variation to each participant to ensure participants encountered all chatbot variations (faceless avatar, static human-like avatar with dynamic expressions) with equal probability.

The questionnaire included several key sections. First, an initial set of demographic questions gathered background information on participants, including age, gender, ICT knowledge, and familiarity with chatbots. Next, an introductory section presented participants with a brief description of the customer support scenario. Participants were instructed to imagine themselves as the customer experiencing this situation.

Following the scenario introduction, participants viewed a pre-recorded animation depicting the customer interacting with the chatbot. A timer measured viewing time to assess if participants watched the entire video before proceeding with the questionnaire. Finally, the survey concluded with a chatbot evaluation section. In this section, participants answered multiple Likert-scale questions evaluating various aspects of the chatbot.

A total of 93 responses were collected. However, only 73 responses were considered complete and unique. Incomplete surveys or those from participants who didn't watch the entire video were excluded.

Table 1 shows the survey's item's, alongside their respective sources, model variables and format.

Variable	Question	Choices	Source
General questions			
Age	What is your age?	Numeric input	Standard demographic question
Gender	What is your gender?	Male/Female/Prefer not to say	Standard demographic question
ICT	Do you work within the ICT? (Information and Communications Technology) sector?	Yes/No	Standard demographic question
Participants are shown 1 of 3 chatbots at random			
Human Likeness questions (Now that you have seen the conversation, please answer the following 6 questions about the appearance of the chatbot Eva.)			
Human	I found the chatbot to be:	Machine-Like – Human-Like	Adapted from Araujo (2018)
Natural	I found the chatbot to be:	Unnatural – Natural	Adapted from Araujo (2018)
Lifelike	I found the chatbot to be:	Artificial – Lifelike	Adapted from Araujo (2018)
Appearance	How much did you like the appearance of the chatbot?	Dislike a great deal – Like a great deal	Adapted from Bartneck et al. (2008)
Importance of Appearance	I find a chatbot's appearance to be:	Not at all important – Extremely important	Moderator question
Professional	The chatbot's appearance came across as:	Extremely unprofessional – Extremely professional	Adapted from Corritore et al. (2005)

Variable	Question	Choices	Source
Statements regarding satisfaction (The following 3 questions are about your user experience with the chatbot Eva.)			
Satisfaction	I would be satisfied with the chatbot.	Strongly disagree – Strongly agree	Adapted from Kvale et al. (2021a)
Trust	The chatbot is trustworthy.	Strongly disagree – Strongly agree	Adapted from Corritore et al. (2005)
Reuse Intention	I would use this chatbot again if I were to have a similar problem.	Definitely will not – Definitely will	Adapted from Venkatesh et al. (2012)
Personality Questions (In the next 5 questions there are words which can be used to describe one's personality. Select the option which you think fits your personality best ranging from Strongly disagree to Strongly agree.)			
P1	Agreeable, Kind	Strongly disagree – Strongly agree	Gosling et al. (2003)
P2	Dependable, Organized	Strongly disagree – Strongly agree	Gosling et al. (2003)
P3	Emotionally stable, Calm	Strongly disagree – Strongly agree	Gosling et al. (2003)
P4	Open to experience, Imaginative	Strongly disagree – Strongly agree	Gosling et al. (2003)
P5	Extraverted, Enthusiastic	Strongly disagree – Strongly agree	Gosling et al. (2003)

Table 1. Questionnaire items.

4.2.1. Sample characteristics.

The survey received 73 complete responses. Participants were comprised of 42 males (57.5%) and 31 females (42.5%). In terms of Information, Communication, and Technology (ICT) knowledge, 22 participants (30.1%) reported working in the ICT sector, while 51 (69.9%) did not.

The age of participants ranged from 18 to 84 years old, with an average age of 39.6 years. The distribution of ages exhibited two central areas, with a larger group between 18 and 30 years old and another between 50 and 60 years old.

The histogram in Figure 5 provides an overview of the age distribution between the age participants in the study. The figure highlights the central tendency and the variability of the age data.

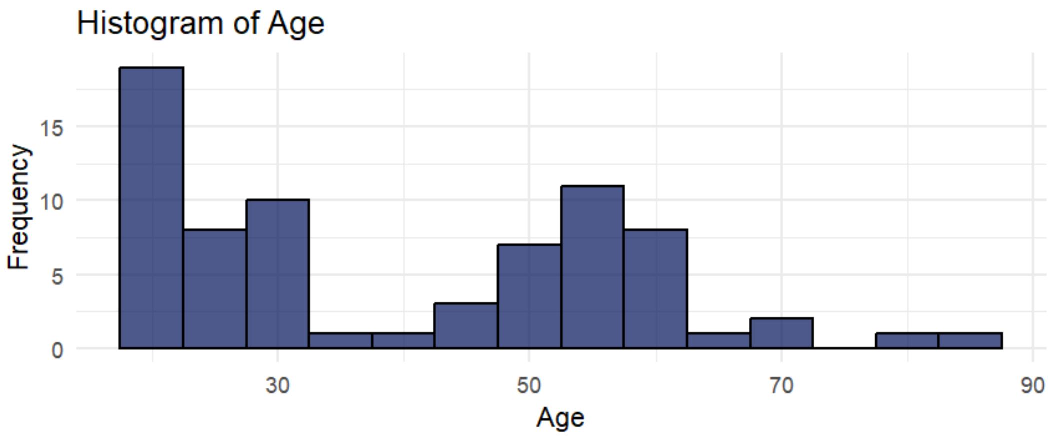

Figure 5. Histogram which shows the age distribution of the survey by comparing age and the frequency.

4.3. Data analysis

To address the research objectives, different statistical techniques were used to analyze the collected survey data. The first research goal was to investigate the influence of chatbot design variations (faceless avatar, static human-like avatar, human-like avatar with dynamic expressions) on perceived Social Presence dimensions (Human-likeness, Naturalness, etc.) Secondly, we sought to understand how these Social Presence dimensions impact Consumer Satisfaction, Trust, and Intention to reuse customer support chatbots.

The analysis was based on a two-step approach. First, an Analysis of Variance (ANOVA) test with a post-hoc Tukey's Honest Significant Difference (HSD) test was conducted to determine if significant differences existed in consumer ratings of Social Presence across the three chatbot variations. Second, Cumulative Link Models (CLMs) were used to model the relationships between the Social Presence dimensions and the response variables (Satisfaction, Trust, and Intention to reuse).

4.3.1. ANOVA and Tukey's HSD tests

An ANOVA test was conducted to check for the presence of statistically significant differences between the Social Presence dimensions (*Human Likeness*, *Naturalness*, *Lifelikeness*, *Professionalism,* and *Appearance*) across distinct chatbot variants. ANOVA is appropriate for this analysis because it allows for simultaneous comparisons between the three chatbot types (faceless avatar, static human-like avatar, human-like avatar with dynamic expressions).

In R software, the ANOVA test was implemented using the aov() function (Anova Function – Rdocumentation), with chatbot type as the independent variable.

Following the ANOVA test, post-hoc testing was necessary to assess the significance of the differences between the chatbot types. Tukey's HSD test was selected for this task since it compares all possible pairs of means. The Tukey's HSD test was implemented using the TukeyHSD() function in R (TukeyHSD Function – Rdocumentation).

4.3.2. Cumulative Link Model Analysis and Likelihood Ratio

The survey data consisted of Likert-scale responses. Since Likert-scale data is ordinal (structured into ranked categories), Cumulative Link Models (CLMs) were chosen to conduct the analysis. CLMs are appropriate for analyzing ordinal data such as star ratings (Agresti, 2010). In this study, five-point Likert-scales are used, which are similar to star ratings. Therefore, the CLM model is a good fit for the data. CLMs are available in the *R statistical software's ordinal Christensen & Brockhoff, 2013) package*.

A critical step in CLM analysis is selecting the link function that best describes the relationship between the ordinal responses and the predictor variables. Several link functions (logit, probit, cloglog, loglog, and cauchit) were evaluated to determine the best fit based on the Akaike Information Criterion (AIC) score.

For each of the three dependent variables (satisfaction, trust, and intention to use again), a separate CLM analysis was conducted, and the model with the lowest AIC score was chosen for further analysis in each case. After determining the AIC scores of the several link functions for each model (Satisfaction, Trust, and Reuse Intention), the best fitting link functions were chosen. In the case of Satisfaction and Reuse Intention *cauchit* was the best fit with an AIC of 155.53 and 119.42, respectively. For the response variable Trust, the best fitting link function was *cloglog* with an AIC of 182.68.

The chosen link functions were then used to estimate the parameters of the respective CLMs. The models explored how independent variables (Human Likeness, Naturalness, Lifelikeness, Professionalism, and Appearance) influenced each of the three dependent response variables. Additionally, the models included an interaction term (Appearance*Importance of Appearance), and several control variables representing participant characteristics (Gender, Age, ICT knowledge, Personality items P1-P5).

Equation 1 presents the CLM formulation used for the analysis. Y_i represents the response variable used in each model: *Satisfaction*, *Trust*, and *Intention to Use Again*. For each response variable, the independent variables are modeled. These include *Human Likeness*, *Naturalness*, *Lifelikeness*, *Professionalism* and *Appearance*. *Appearance* is included along with its interaction term, the Importance *of Appearance*. The dummy variables *Gender*, *Age*, *ICT*, *P1-P5*, *Human Persona*, and *Emotions* are modeled as $\alpha_{i.}$ The variable *Human Persona* indicates whether the chatbot had a Human avatar or if it was the chatbot without an avatar. The variable *Emotions* indicates whether the chatbot had facial expressions or if it was one of the other two chatbots.

$$LF\big(P(Y_{\mathrm{i}} \le j)\big) = \Theta_{\mathrm{j}} + ß_1 * Human\ Likeness_{\mathrm{i}} + \ + ß_2 * Naturalness_{\mathrm{i}} + ß_3 * Lifelikeness_{\mathrm{i}} \\ + ß_4 * Professionalism_{\mathrm{i}} + ß_5 * \ (Appearance_{\mathrm{i}} * Importance_{\mathrm{i}}) + \\ \alpha_{\mathrm{i}} - Importance_{\mathrm{i}}$$

Equation 1. Cumulative Link Model formulation.

LF represents the chosen link functions, *cauchit* for Satisfaction and Reuse intention, *cloglog*, for Trust. $P(\mathrm{Yi} \le j)$ is the cumulative probability of the response variable Y_i being less than or equal to the category *j*. The *j* represents the cut points between categories. Θ_j are the thresholds for the cumulative probabilities. The *'i'* represents each observation in the dataset ranging from *i = 1 to 73*.

Following the construction of the Cumulative Link Models for Satisfaction, Trust, and Reuse Intention, post-hoc testing was conducted via the Likelihood ratio test. For each of the three CLM models a null model was made, this model contained only the control variables. The models are then compared to their null models using a likelihood ratio test (R: Likelihood Ratio Test), which follows a chi-squared distribution, to determine if the models were statistically significant. The results indicate that all models have significantly better AIC scores compared to their respective null models (p <0.001).

5. Results

This section presents the findings from the analysis conducted using the ANOVA test and the Cumulative Link Model (CLM). Each hypothesis will be reiterated along with the corresponding decisions. The results are organized into two parts: (I) The Influence of Social Presence on Satisfaction, Trust, and Intention to Reuse.

5.1. Influence of chatbot design on Social Presence dimensions

As outlined in the methodology section, an ANOVA test was conducted to test Hypotheses of H1, H2, and H3, comparing the three chatbot scores. The results of the ANOVA test indicated that there were no significant differences in the scores among the three chatbot types. While the data showed some variation in ratings, such as the average Trust ratings where the chatbot with facial expressions received an average rating of 3.53 compared to 3 for the other two types, these differences were not statistically significant. This conclusion was supported by post-hoc analysis using Tukey's HSD test. Consequently, Hypotheses H1, H2, and H3 were rejected. The hypotheses and their corresponding decisions are summarized in Table 2.

Hypothesis	Decision
H1a: Perceived Human-likeness will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.	Reject
H1b: Perceived Naturalness will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.	Reject
H1c: Perceived Lifelikeness will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.	Reject
H1d: Perceived Professionalism will be significantly higher in chatbots with a (static) human-like avatar compared to chatbots without an avatar.	Reject
H2a: Perceived Human-likeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.	Reject
H2b: Perceived Naturalness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.	Reject

H2c: Perceived Lifelikeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.	Reject
H2d: Perceived Professionalism will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots without an avatar.	Reject
H3a: Perceived Human-likeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.	Reject
H3b: Perceived Naturalness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.	Reject
H3c: Perceived Lifelikeness will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.	Reject
H3d: Perceived Professionalism will be significantly higher in chatbots with a human-like avatar featuring dynamic facial expressions compared to chatbots with a static human-like avatar.	Reject

Table 2. Overview of hypotheses H1, H2, and H3 and their decisions after performing ANOVA and Tukey's HSD tests.

5.2. Impact of Social Presence on Satisfaction, Trust, and Reuse Intention

5.2.1. H4 Impact of Social Presence on Satisfaction

The Cumulative Link Model of Satisfaction highlighted multiple significant relationships. The model returned significant results for Human-likeness, Naturalness, and Professionalism, which are highlighted in Table 3. The estimates of these variables were all positive, which suggests that a higher rating for one of these variables leads to a higher Satisfaction score. This means that three of the four hypotheses of H4 are not rejected (Table 4).

There were also significant results for the Control variables. Personality questions P1, P4, and P5 were indicated to have a significant relationship with Satisfaction. The estimates of Extraversion (P1) and Openness to Experience (P5) were negative, thus indicating that a higher score in Extraversion (P1) or Openness to Experience (P5) will result in a lower Satisfaction score. The estimate of Emotional Stability (P4) had a positive estimate which suggests that a higher score on the Emotional Stability (P4) question will result in a higher Satisfaction score.

H4, Response Variable Satisfaction, Link function: Cauchit	
Variable	Effect (SE)
Main effects	
Human-likeness	**5.221* (2.422)**
Naturalness	**3.874* (1.926)**
Lifelikeness	-1.637 (1.016)
Professionalism	**9.897* (4.495)**
Appearance	-2.574 (1.402)
Control variables	
GenderMale	4.978 (2.955)
Age	-0.193 (0.105)
ICT	3.705 (2.227)
P1; Extraversion	**-3.291* (1.578)**
P2; Agreeableness	1.842 (1.080)
P3; Conscientiousness	-0.870 (0.557)
P4; Emotional Stability	**4.496* (2.028)**
P5; Openness to Experience	**-2.977* (1.383)**
HumanPersona	0.228 (1.215)
Emotions	0.406 (1.734)
Appearance: Importance	0.661 (0.342)

Table 3. Results of the CLM model for the response variable Satisfaction.

Note Significant Variables are marked in ***bold****. *p <0.05; **p < 0.01; *** < 0.001; N = 73*

‘

Hypothesis	Decision
H4a: Perceived Human-likeness in a customer support chatbot will be positively associated with Consumer Satisfaction.	Do not Reject
H4b: Perceived Naturalness in a customer support chatbot will be positively associated with Consumer Satisfaction.	Do not Reject
H4c: Perceived Lifelikeness in a customer support chatbot will be positively associated with Consumer Satisfaction.	Reject
H4d: Perceived Professionalism in a customer support chatbot will be positively associated with Consumer Satisfaction.	Do not Reject

Table 4. Overview of the Hypotheses of H4 and their decisions following the CLM model.

5.2.2. H5 Impact of Social Presence on Trust

The Cumulative Link Model of Trust highlighted multiple significant relationships. The model returned significant results for Human-likeness, Naturalness, and Professionalism, which are highlighted in Table 5. The estimates of these variables were all positive, which suggests that a higher rating for one of these variables leads to a higher Trust score. This means that three of the four hypotheses of H5 are not rejected (Table 6).

Additionally, the Control variables returned some significant results. Gender Male was reported as significant and positively correlated with trust. This suggests that if respondents are of the Male gender, this influences the Trust score positively. Not only Gender proved to play a significant role within this model, as Age was reported with $p < 0.01$ (**). The estimate of Age is on the smaller side in comparison ton the other estimates, with only 0.038 compared to the other estimates which are larger than 0.4. This could suggest that when a respondent is older the Trust score is impacted positively, with this positive effect being very minimal. Agreeableness (P2) and Conscientiousness (P3) also returned significant results at $p < 0.05$ with Agreeableness (P2) having a positive estimate whilst Conscientiousness (P3) has a similar negative estimate.

H5, Response Variable Trust, Link function: Cloglog	
Variable	Effect (SE)
Main effects	
Human-likeness	**0.569* (0.290)**
Naturalness	**0.598* (0.292)**
Lifelikeness	-0.157 (-0.595)
Professionalism	**0.641* (0.268)**
Appearance	0.382 (0.324)
Control variables	
GenderMale	**0.806* (0.366)**
Age	**0.038** (0.013)**
ICT	-0.640 (0.415)
P1; Extraversion	-0.056 (0.214)
P2; Agreeableness	**0.425* (0.176)**
P3; Conscientiousness	**-0.404* (0.161)**
P4; Emotional Stability	-0.210 (0.161)
P5; Openness to Experience	0.096 (0.137)
HumanPersona	0.155 (0.420)
Emotions	0.893 (0.474)
Appearance: Importance	0.025 (0.057)

Table 5. Results of the CLM model for the response variable Trust.

Note Significant Variables are marked in ***bold****. *p <0.05; **p < 0.01; *** < 0.001; N = 73*

Hypothesis	Decision
H5a: Perceived Human-likeness in a customer support chatbot will be positively associated with Trust.	Do not Reject
H5b: Perceived Naturalness in a customer support chatbot will be positively associated with Trust.	Do not Reject
H5c: Perceived Lifelikeness in a customer support chatbot will be positively associated with Trust.	Reject
H5d: Perceived Professionalism in a customer support chatbot will be positively associated with Trust.	Do not Reject

Table 6. Overview of the Hypotheses of H5 and their decisions following the CLM model.

5.2.3. H6 Impact of Social Presence on Reuse Intention

The Cumulative Link Model of Reuse intention indicated only one significant relationship. The four social presence dimensions did not yield significant results. The results are highlighted in Table 7. Consequently, the hypotheses of H6 highlighted in Table 8 were rejected.

Additionally, most control variables reported no significant correlations. However, there was one exception: the control variable Emotions, showed a significant result with $p < 0.05$ with an estimate of 8.7. This variable shows whether a chatbot had facial expressions or not. Since the estimate is positive, this suggests that Emotions positively influence Reuse intention.

H5, Response Variable UseAgain, Link function: Cauchit	
Variable	Effect (SE)
Main effects	
Human-likeness	0.404 (0.984)
Naturalness	8.048 (4.180)
Lifelikeness	-2.126 (-1.168)
Professionalism	3.561 (2.291)
Appearance	0.135 (1.087)
Control variables	
GenderMale	-0.439 (1.334)
Age	-0.190 (0.106)
ICT	8.789 (5.050)
P1; Extraversion	-1.690 (1.470)
P2; Agreeableness	2.113 (1.296)
P3; Conscientiousness	-0.560 (0.685)
P4; Emotional Stability	1.180 (0.978)
P5; Openness to Experience	1.072 (0.768)
HumanPersona	0.577 (1.911)
Emotions	**8.702* (4.325)**
Appearance: Importance	-0.098 (0.295)

Table 7. Results of the CLM model for the response variable Reuse Intention.

*Note Significant Variables are marked in **bold**. *p <0.05; **p < 0.01; *** < 0.001; N = 73*

Hypothesis	Decision
H6a: Perceived Human-likeness in a customer support chatbot will be positively associated with Reuse Intention.	Reject
H6b: Perceived Naturalness in a customer support chatbot will be positively associated with Reuse Intention.	Reject
H6c: Perceived Lifelikeness in a customer support chatbot will be positively associated with Reuse Intention.	Reject
H6d: Perceived Professionalism in a customer support chatbot will be positively associated with Reuse Intention.	Reject

Table 8. Overview of the Hypotheses of H6 and their decisions following the CLM model.

6. Discussion and Conclusion

This study aimed to explore the influence of chatbot avatar design on Social Presence. With the constant evolution of chatbots, the study investigated the impact of chatbot dynamic avatar appearances in a business context. Additionally, the study tested the impact of Social Presence dimensions, namely Human-likeness, Naturalness, Lifelikeness, and Professionalism, on customer Satisfaction, Trust, and Reuse Intention. By integrating these findings, the study seeks to contribute to the enhancement of chatbot design techniques in business settings.

6.1. Theoretical Implications

This study contributes to the existing literature by investigating the visual design and appearance of chatbots, as opposed to focusing on text-based anthropomorphic design. Araujo (2018) explored the effects of interacting with disembodied agents through text, examining customer outcomes such as attitudes, satisfaction, and emotional connection towards the company. Similarly, Konya-Baumbach et al. (2023) assessed the effectiveness of chatbot anthropomorphism using human-like linguistic cues, finding positive effects on satisfaction, trust, and other factors such as purchase intention and word-of-mouth, driven by Social Presence.

In contrast to Araujo (2018) and Konya-Baumbach et al. (2023), this study focused on the impact of visual design, specifically dynamic facial expressions, on Social Presence, Satisfaction, Trust, and Reuse Intention. The findings suggest that while dynamic facial expressions do not significantly enhance Social Presence dimensions compared to static avatars, they do positively impact Reuse Intention scores. The study also shows that Satisfaction and Trust are positively impacted by the perceived Human-likeness, Naturalness, and Professionalism of the chatbot. Besides these main effects, there were also some significant results within the control variables. For all five of the Personality questions, there was a significant impact on either Satisfaction or Trust, whilst Reuse Intention was not significantly impacted by these questions. Trust scores were also positively impacted by age and by the Male Gender. These results indicate that the personality of users influences the Satisfaction and Trust scores of users. Besides personalities, the results indicate that Male users' Trust in the chatbot in this study is significantly higher than that of Female users.

These results further expand on the impact of Social Presence on Chatbot Design found in previous studies. They highlight the role of visual design in chatbot interactions, suggesting that different elements of Social Presence may influence various user experience outcomes. The results also highlight that users' personalities may influence how they perceive chatbots.

6.2. Practical Implications

For this study, it was hypothesized that a chatbot with changing facial expressions would score significantly higher than one with only a human avatar or one without an avatar. On some of the questions, there was some variation between the scores of the chatbot types. The results of the tests, however, show that there is no significant difference between the scores of the three chatbot types. The study was unable to prove that making a change towards changing facial expressions will yield significantly better scores of Social Presence dimensions. In practice, a Human-like or No avatar will perform similarly to a chatbot with changing facial expressions on these dimensions.

The study found that certain Social Presence dimensions have a significant impact on users' Satisfaction and Trust scores, namely Human-likeness, Naturalness, and Professionalism had a significant positive impact on these scores. These three dimensions have no significant impact on Reuse Intention. One of the four dimensions discussed in this study, lifelikeness, had no significant impact on users' Satisfaction, Trust, and Reuse Intention scores. This indicates that chatbot practitioners should aim to enhance their chatbot's Human-likeness, Naturalness, and Professionalism to get higher Satisfaction, Trust scores from their users.

For Reuse Intention the only significant variable was those of Emotions. This indicates that when the users used the chatbot with dynamic facial expressions, this positively impacted the Reuse Intention score. When taking these results into practice, this means that using a chatbot with facial expressions could lead to better Reuse Intention scores among users.

In conclusion, chatbot practitioners should consider the positive impact of three of the Social Presence Dimensions, Human-likeness, Naturalness, and Professionalism on the Satisfaction and Trust scores of users. Furthermore, practitioners will have to consider the role dynamic facial expressions play within the user's Reuse Intention score. Using chatbots with dynamic facial expressions could result in higher Reuse Intention scores. The Social Presence Dimension of Lifelikeness did not report a significant impact on users' Satisfaction, Trust, and Reuse Intention scores. Age only had a small impact, although significant, on Trust. The personality questions showed some significance, but this study did not review this in depth. More on this can be found in the next section.

6.3. Limitations and Opportunities for Future Research

It is important to recognize that this research has several limitations, which could present opportunities for future research.

Firstly, the study used chatbot animations instead of using a fully functioning chatbot interface. As previously mentioned, this choice was made due to the time constraints of the research and to isolate the effect of the chatbot avatar, ensuring the only difference between the three chatbots was the avatar. Future research could benefit from developing a real chatbot with dynamic expressions. This would allow users to interact with the chatbot in unique ways, which better imitates the way chatbots are used in business settings.

Secondly, the survey population was small, with the data of 73 respondents which could be used for the CLM. The sample sizes for the three chatbot types were 26,24, and 23 for the dynamic, human, and nonhuman chatbot types in the ANOVA. Future studies could validate these models and analyze them with a larger survey population to improve the reliability and generalizations of the findings.

Thirdly, this study explored the effect of Social Presence dimensions in a positive business scenario where the chatbot can help the chatbot user. It would be valuable to investigate the impact of these dimensions on Satisfaction, Trust, and Reuse Intention when the chatbot is unable to help the user. Besides the impact of the Social Presence dimensions, it could then also be explored if there is a difference between the three chatbots in this negative scenario.

Lastly, due to time and length constraints, the study did not further explore the impact of personality dimensions on Satisfaction and Trust scores. Since all five of the personality questions seem to have an impact on either Satisfaction or Trust scores it could be interesting to further study how users' personalities influence these scores. Chatbot practitioners should consider that users' personalities can influence the Satisfaction and Trust scores both positively and negatively.

Acknowledgments

I would like to thank my supervisor Dr. Natalia Amat Lefort for her continued support and encouragement throughout the research project. Her insights and contributions have been of great worth and have helped shape this thesis as is. I could not have completed this research without her.

I also wish to thank my second supervisor Dr. Yingjie Fan for supporting both me and Dr. Amat Lefort during the project. Her flexibility, support, and feedback have been of great help in the development of the thesis.

At last, I would also like to thank Dr. Marieke Vinkenoog for her assistance during the statistical analysis part of the thesis. Her help regarding which models and statistical tests to use given the research model was of great importance.

References

- Adam, M., Wessel, M., & Benlian, A. (2020). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7
- Adamopoulou, E., & Moussiades. (2020). An overview of Chatbot Technology. in IFIP International Conference on Artificial Intelligence Applications and Innovations. *Springer, Cham.*, (pp. 373-383).
- Agarwal, R., & Wadhwa, M. (2020). Review of State-of-the-Art Design Techniques for Chatbots. *SN Computer Science*, *1*(5). https://doi.org/10.1007/s42979-020-00255-3
- Agresti, A. (2010). Analysis of Ordinal Categorical Data. In *Wiley series in probability and statistics*. https://doi.org/10.1002/9780470594001
- Albrecht, T., Rausch, T. M., & Derra, N. D. (2021). Call me maybe: Methods and practical implementation of artificial intelligence in call center arrivals' forecasting. *Journal Of Business Research*, *123*, 267–278. https://doi.org/10.1016/j.jbusres.2020.09.033
- Aljazzaf, Z. M., Perry, M., & Capretz, M. A. (2010). Online Trust: Definition and Principles. *In Proceedings Of The ICCGI'10: Proceedings Of The 2010 Fifth International Multi-Conference On Computing in The Global Information Technology*. https://doi.org/10.1109/iccgi.2010.17
- Androutsopoulou, A., Karacapilidis, N., Loukis, E. N., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, *36*(2), 358–367. https://doi.org/10.1016/j.giq.2018.10.001
- *anova function - RDocumentation*. (z.d.). https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/anova
- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051
- Atiyah, A., Jusoh, S., & Alghanim, F. (2019). Evaluation of the Naturalness of Chatbot Applications. *2019 IEEE Jordan International Joint Conference On Electrical Engineering And Information Technology (JEEIT)*. https://doi.org/10.1109/jeeit.2019.8717455
- Baabdullah, A. M., Alalwan, A. A., Algharabat, R., Metri, B. A., & Rana, N. P. (2022). Virtual agents and flow experience: An empirical examination of AI-powered chatbots. *Technological Forecasting And Social Change*, *181*, 121772. https://doi.org/10.1016/j.techfore.2022.121772
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2008). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal Of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3
- Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence*, *12*(5), 456–480. https://doi.org/10.1162/105474603322761270

- Brandtzæg, P. B., & Fölstad, A. (2018). Chatbots. *Interactions*, *25*(5), 38–43. https://doi.org/10.1145/3236669
- Chatbot Research and Design. (2020). In *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-030-39540-7
- Chaves, A. P., & Gerosa, M. A. (2020). How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal Of Human-Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438
- Choi, S. W., Mattila, A. S., & Bolton, L. E. (2020). To Err Is Human(-oid): How Do Consumers React to Robot Service Failure and Recovery? *Journal Of Service Research*, *24*(3), 354–371. https://doi.org/10.1177/1094670520978798
- Christensen, R. H. B., & Brockhoff, P. B. (2013). Analysis of sensory ratings data with cumulative link models. *Journal de La Société Française de Statistique & Revue de Statistique Appliquée*, *154*(3), 58–79. https://orbit.dtu.dk/files/79012845/Christensen%20and%20Brockhoff%202013.pdf
- Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International Journal Of Human-computer Studies*, *58*(6), 737–758. https://doi.org/10.1016/s1071-5819(03)00041-7
- Corritore, C. L., Marble, R. P., Wiedenbeck, S., Kracher, B., & Essential Products, Inc. (2005). Measuring Online Trust of Websites: Credibility, Perceived Ease of Use, and Risk. In *AMCIS 2005 Proceedings* [Conference-proceeding]. https://aisel.aisnet.org/amcis2005/370/
- Daft, R. L., & Lengel, R. H. (1984). Information Richness: A new approach to managerial decision making. *Academy Of Management Review*, 9(2), 199-213.
- Danaher, P. J., & Haddrell, V. E. (1996). A comparison of question scales used for measuring customer satisfaction. *International Journal Of Service Industry Management*, *7*(4), 4–26. https://doi.org/10.1108/09564239610129922
- Dekkal, M., Arcand, M., Tep, S. P., Rajaobelina, L., & Ricard, L. (2023). Factors affecting user trust and intention in adopting chatbots: the moderating role of technology anxiety in insurtech. *Journal Of Financial Services Marketing*. https://doi.org/10.1057/s41264-023-00230-y
- Drift, & Drift. (2023, 14 april). The State of Chatbots Report: How Chatbots Are Reshaping Online Experiences. *Drift*. https://www.drift.com/blog/Chatbots-report/
- Følstad, A., & Brandtzæg, P. B. (2020). Users' experiences with chatbots: findings from a questionnaire study. *Quality And User Experience*, *5*(1). https://doi.org/10.1007/s41233-020-00033-2
- Følstad, A., Kvale, K., & Haugstveit, I. M. (2014). Customer support as a source of usability insight. *NordiCHI '14: Proceedings Of The 8th Nordic Conference On Human-Computer Interaction: Fun, Fast, Foundational*. https://doi.org/10.1145/2639189.2639232
- Følstad, A., & Skjuve, M. (2019). Chatbots for customer service. *CUI '19: Proceedings Of The 1st International Conference On Conversational User Interfaces*. https://doi.org/10.1145/3342775.3342784

- Gartner. (2022, 27 juli). *Gartner Predicts Chatbots Will Become a Primary Customer Service Channel Within Five Years*. www.gartner.com. Geraadpleegd op 10 april 2024, van https://www.gartner.com/en/newsroom/press-releases/2022-07-27-gartner-predicts-chatbots-will-become-a-primary-customer-service-channel-within-five-years
- Goodwin, C., & Smith, K. L. (1990). Courtesy and Friendliness: Conflicting Goals for the Service Providers. *Journal Of Services Marketing/The Journal Of Services Marketing*, *4*(1), 5–20. https://doi.org/10.1108/eum0000000002500
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal Of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/s0092-6566(03)00046-1
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254
- Haugeland, I. K. F., Følstad, A., Taylor, C., & Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. *International Journal Of Human-Computer Studies*, *161*, 102788. https://doi.org/10.1016/j.ijhcs.2022.102788
- Hess, U. (2020). Body language. In *Springer eBooks* (pp. 527–531). https://doi.org/10.1007/978-3-319-24612-3_647
- Jain, M., Kumar, P., Kota, R., & Patel, S. (2018). Evaluating and Informing the Design of Chatbots. *DIS '18: Proceedings Of The 2018 Designing Interactive Systems Conference*. https://doi.org/10.1145/3196709.3196735
- Khwaja, A., & Yang, N. (2022). Quantifying the link between employee engagement, and customer satisfaction and retention in the car rental industry. *Quantitative Marketing And Economics*, *20*(3), 275–292. https://doi.org/10.1007/s11129-022-09253-6
- Kim, N. (2019). Understanding Key Antecedents of Consumer Loyalty toward Sharing-Economy Platforms: The Case of Airbnb. *Sustainability*, *11*(19), 5195. https://doi.org/10.3390/su11195195
- Klein, K., & Martinez, L. F. (2022). The impact of anthropomorphism on customer satisfaction in chatbot commerce: an experimental study in the food sector. *Electronic Commerce Research*, *23*(4), 2789–2825. https://doi.org/10.1007/s10660-022-09562-8
- Konya-Baumbach, E., Biller, M., & Von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior*, *139*, 107513. https://doi.org/10.1016/j.chb.2022.107513
- Kvale, K., Freddi, E., Hodnebrog, S., Sell, O. A., & Følstad, A. (2021). Understanding the User Experience of Customer Service Chatbots: What Can We Learn from Customer Satisfaction Surveys? In *Lecture Notes in Computer Science* (pp. 205–218). https://doi.org/10.1007/978-3-030-68288-0_14
- Kwangsawad, A., & Jattamart, A. (2022). Overcoming customer innovation resistance to the sustainable adoption of chatbot services: A community-enterprise perspective in Thailand. *Journal Of Innovation & Knowledge/Journal Of Innovation And Knowledge*, *7*(3), 100211. https://doi.org/10.1016/j.jik.2022.100211
- *lifelike*. (2024). https://dictionary.cambridge.org/dictionary/english/lifelike

- Luger, E., & Sellen, A. (2016). “Like Having a Really Bad PA”. *CHI ’16: Proceedings Of The 2016 CHI Conference On Human Factors in Computing Systems*. https://doi.org/10.1145/2858036.2858288
- Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *EM*, *32*(4), 1993–2020. https://doi.org/10.1007/s12525-022-00605-4
- McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa . . . examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28–37. https://doi.org/10.1016/j.chb.2019.05.009
- *Measuring patient satisfaction: how to do it and why to bother*. (1999, 1 januari). PubMed. https://pubmed.ncbi.nlm.nih.gov/10344931
- Meyer, J., Miller, C., Hancock, P. A., De Visser, E. J., & Dorneich, M. C. (2016). Politeness in Machine-Human and Human-Human Interaction. *Proceedings Of The Human Factors And Ergonomics Society Annual Meeting/Proceedings Of The Human Factors And Ergonomics Society . . . Annual Meeting*, *60*(1), 279–283. https://doi.org/10.1177/1541931213601064
- Mimoun, M. S. B., Poncin, I., & Garnier, M. (2017). Animated conversational agents and e-consumer productivity: The roles of agents and individual characteristics. *Information & Management*, *54*(5), 545–559. https://doi.org/10.1016/j.im.2016.11.008
- Mittal, V., Han, K., & Westbrook, R. A. (2017). Customer Engagement and Employee Engagement: A Research Review and Agenda. In *Springer eBooks* (pp. 173–201). https://doi.org/10.1007/978-3-319-61985-9_8
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal Of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153
- naturalness. (2024). In *dictionary.cambridge.org*. https://dictionary.cambridge.org/dictionary/english/naturalness
- Oliver, R. L. (1999). Whence consumer loyalty? *Journal Of Marketing*, *63*, 33. https://doi.org/10.2307/1252099
- Otto, A. S., Szymanski, D. M., & Varadarajan, R. (2019). Customer satisfaction and firm performance: insights from over a quarter century of empirical research. *Journal Of The Academy Of Marketing Science*, *48*(3), 543–564. https://doi.org/10.1007/s11747-019-00657-7
- Petty, R. E., & Cacioppo. (1986). The elaboration likelihood model of persuasion. *Communication And Persuasion*, 1-24.
- *professionalism*. (2024). https://dictionary.cambridge.org/dictionary/english/professionalism
- *R: Likelihood ratio test*. (z.d.). https://search.r-project.org/CRAN/refmans/epiDisplay/html/lrtest.html#:~:text=Likelihood%20ratio%20test%20checks%20the%20difference%20between%20-2%2AlogLikelihood,dataset%20for%20both%20models%20MUST%20be%20the%20same.

- Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the Impact of "Humanizing" Customer Service Chatbots. *Information Systems Research*, *32*(3), 736–751. https://doi.org/10.1287/isre.2021.1015
- Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal Of Business Research*, *115*, 14–24. https://doi.org/10.1016/j.jbusres.2020.04.030
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. https://ci.nii.ac.jp/ncid/BA22930972
- Silva, F. A., Shojaei, A. S., & Barbosa, B. (2023). Chatbot-Based Services: A Study on Customers' Reuse Intention. *Journal Of Theoretical And Applied Electronic Commerce Research*, *18*(1), 457–474. https://doi.org/10.3390/jtaer18010024
- Smestad, T. L. (2018). *Personality Matters! Improving The User Experience of Chatbot Interfaces - Personality provides a stable pattern to guide the design and behaviour of conversational agents*. https://brage.bibsys.no/xmlui/handle/11250/2502575
- Toader, D., Boca, G. D., Toader, R., Măcelaru, M., Toader, C., Ighian, D., & Radulescu, A. (2019). The Effect of Social Presence and Chatbot Errors on Trust. *Sustainability*, *12*(1), 256. https://doi.org/10.3390/su12010256
- *TukeyHSD function - RDocumentation*. (z.d.). https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/TukeyHSD
- Türkyılmaz, A., & Özkan, C. (2007). Development of a customer satisfaction index model. *Industrial Management + Data Systems/Industrial Management & Data Systems*, *107*(5), 672–687. https://doi.org/10.1108/02635570710750426
- Valério, F. A. M., Guimarães, T. G., Prates, R. O., & Candello, H. (2017). Here's What I Can Do. *IBM*. https://doi.org/10.1145/3160504.3160544
- VectorKIF. (2018, 12 december). *Face expressions of woman with blond hair. Different female emotions*. . . iStock. https://www.istockphoto.com/nl/vector/gezicht-uitingen-van-vrouw-met-blond-haar-gm1081125798-289884774
- Venkatesh, N., Thong, N., & Xu, N. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *Management Information Systems Quarterly*, *36*(1), 157. https://doi.org/10.2307/41410412
- Walther, J. B. (1992). Interpersonal Effects in Computer-Mediated Interaction. *Communication Research*, *19*(1), 52–90. https://doi.org/10.1177/009365092019001003

Appendix

Question	Choices	Source
General questions		
Q1. What is your age?	Numeric input	Standard demographic question
Q2. What is your gender?	Male/Female/Prefer not to say	Standard demographic question
Q3. Are you a Computer Science student	Yes/No	Standard demographic question
Q4. Do you work within the ICT (Information and Communications Technology) sector?	Yes/No	Standard demographic question
Q5. How familiar are you with chatbots?	Very unfamiliar – Very Familiar	Adapted from Mimoun et al. (2017)
Participants are shown 1 of 3 chatbots at random		
Human Likeness questions (Now that you have seen the conversation, please answer the following 6 questions about the appearance of the chatbot Eva.)		
Q6. I found the chatbot to be:	Machine-Like – Human-Like	Adapted from Araujo (2018)
Q7. I found the chatbot to be:	Unnatural – Natural	Adapted from Araujo (2018)
Q8. I found the chatbot to be:	Artificial – Lifelike	Adapted from Araujo (2018)
Q9. How much did you like the appearance of the chatbot?	Dislike a great deal – Like a great deal	Adapted from Bartneck et al. (2008)
Q10. I find a chatbot's appearance to be:	Not at all important – Extremely important	Moderator question
Q11. The chatbot's appearance came across as:	Extremely unprofessional – Extremely professional	Adapted from Corritore et al. (2005)

Question	Choices	Source
Statements regarding satisfaction (The following 7 questions are about your user experience with the chatbot Eva.)		
Q12.1 I would be satisfied with the chatbot.	Strongly disagree – Strongly agree	Adapted from Kvale et al. (2021a)
Q12.2 The chatbot did a good job.	Strongly disagree – Strongly agree	Adapted from Corritore et al. (2005)
Q12.3 The chatbot did what I expected.	Strongly disagree – Strongly agree	Adapted from Corritore et al. (2005)
Q13.1 The chatbot is honest.	Strongly disagree – Strongly agree	Adapted from Corritore et al. (2005)
Q13.2 The chatbot is trustworthy.	Strongly disagree – Strongly agree	Adapted from Corritore et al. (2005)
Q14. Overall, I found my experience with the chatbot to be:	Extremely negative – Extremely positive	Adapted from Bartneck et al. (2008)
Q15. I would use this chatbot again if I were to have a similar problem.	Definitely will not – Definitely will	Adapted from Venkatesh et al. (2012)
Personality Questions (In the next 5 questions there are words which can be used to describe one's personality. Select the option which you think fits your personality best ranging from Strongly disagree to Strongly agree.)		
Q16. Agreeable, Kind	Strongly disagree – Strongly agree	Gosling et al. (2003)
Q17. Dependable, Organized	Strongly disagree – Strongly agree	Gosling et al. (2003)
Q18. Emotionally stable, Calm	Strongly disagree – Strongly agree	Gosling et al. (2003)
Q19. Open to experience, Imaginative	Strongly disagree – Strongly agree	Gosling et al. (2003)
Q20. Extraverted, Enthusiastic	Strongly disagree – Strongly agree	Gosling et al. (2003)
Only participants of Chatbot 1 (with the facial expressions) were shown the following question:		
Q21. Did you notice the changing facial expressions of chatbot Eva during the video?	Yes/No	Video question

Appendix A . All questions asked on the Survey, including those which were not used in the models.