Transparent interactions: The effects of human-like design and chatbot disclosure on trust perception and user experience

Master thesis

Annika Setzmann - s2640384

Faculty of Behavioral, Management and Social Sciences

Communication Science

1st Supervisor: Dr. J. Karreman

2nd Supervisor: Dr. E. van Laar

University of Twente, Enschede

November 14, 2024

Abstract

Background

Chatbots are increasingly replacing human agents in online customer interactions, yet people continue to expect them to be as helpful and competent as real humans. To meet these expectations, chatbots must be perceived as trustworthy, and the user experience should be pleasant and effective. Understanding how chatbot design can influence this is therefore essential for optimizing customer service. Human-like design features could foster trust by mimicking natural interaction patterns, but their overall impact on user experience remains underexplored. As AI advancements make chatbots behave increasingly human-like, it becomes vital to understand how transparency about their artificial nature influences customer perceptions. Without such transparency, users may struggle to distinguish whether they are interacting with a bot or a human.

Objectives

Thus, this study looks into how the human-like visual and verbal characteristics of chatbots, along with explicit transparency about their artificial nature, influence user trust and experience. Additionally, it examines whether consistency between a chatbot's human-like appearance and language use plays a role in enhancing trust and user experience.

Methods

A 2x2x2 factorial design was employed, where 192 participants viewed pre-recorded chatbot interactions that varied in terms of visual appearance (human-like or not), verbal communication style (human-like or not), and transparency regarding the chatbot's artificial identity. Participants then provided feedback through a survey on their perceptions of trust and user experience.

Results

The findings show that human-like visual design did not have a significant effect on any aspects of user trust or experience. In contrast, human-like verbal design improved several user experience factors, including attractiveness, efficiency, dependability, and stimulation. It also improved the trust dimension of perceived benevolence. Transparency about the chatbot's artificial nature increased trust, particularly in terms of perceived ability and benevolence, though it did not affect perceived integrity. Additionally, no significant interactions were found between visual and verbal design or between transparency and the visual and verbal design elements.

Conclusion

The findings suggest that while human-like visual design may not play a crucial role, focusing on human-like verbal communication and transparency can significantly enhance user trust and experience. This provides practical guidance on how to effectively combine human-like conversational elements with transparency about the chatbot's artificial nature, to foster trust and improve users' experiences with the chatbot.

Keywords: Human-computer interaction, chatbot, transparency, humanness, trust, user experience, design.

Transparent interactions: The effects of human-like design and chatbot disclosure on trust perception and user experience

As technology keeps improving, the widespread integration of chatbots in service sectors has transformed customer service and business-customer interactions. These AI-powered conversational agents are reshaping how businesses provide services, altering consumer expectations and operational efficiency. Chatbots offer companies significant advantages, including round-the-clock availability, managing multiple inquiries at the same time, and lowering costs by automating routine tasks (Anaya et al., 2024; Xu et al., 2017). Despite this, some companies remain hesitant to fully embrace chatbot technology due to concerns that customers may perceive the user experience as less satisfactory compared to human interaction (Ashfaq et al., 2020). However, chatbots can now be designed to closely resemble human looks and mimic human behavior, making it essential to explore how this level of human-likeness influences user experiences.

The success of chatbots in customer service depends heavily on psychological factors that influence how humans interact with machines (Nadarzynski et al., 2019; Nass et al., 1994). Among these factors, trust plays a central role as it is essential for any meaningful interaction. For customers to accept advice or assistance from a service chatbot, they need to trust that it will act in their best interest, demonstrate competence, and behave reliably—similar to the expectations they would have of a human service agent (Choudhury & Shamszare, 2023; Mayer et al., 1995). Since users often treat chatbots as if they were real people (Kühne & Peter, 2022; Reeves & Nass, 1996), understanding how trust is built in these interactions and identifying the design elements that effectively foster trust are critical for enhancing chatbot success.

Additionally, chatbots are sometimes designed to be so lifelike that users may not immediately realize they are interacting with an AI rather than a human. If this is not clearly disclosed, customers

may feel misled, potentially damaging the company's reputation (Crockett et al., 2019). This is where the concept of transparency becomes essential. Openly acknowledging a chatbot's artificial nature has been shown to enhance its perceived social intelligence, reduce feelings of unease, and increase user affinity (Xu et al., 2023). As a result, transparency may also foster greater trust in chatbots, and ultimately improve the overall customer experience.

Prior research has explored human-like visual and verbal design cues as combined factors (Araujo, 2018; Chen et al., 2023; Chen et al., 2024). Building on this, this study examined visual and verbal human-like design as distinct factors, along with transparency, to understand how each element interacts to shape user trust and experience. Moreover, this research aimed to uncover whether consistency between human-like appearance and speech amplifies trust and enhances user experience—a relationship that has been suggested but not thoroughly examined in earlier work (Miao et al., 2021). Thus, the study was designed to answer the following research questions:

RQ1: *To what extent do visual design, verbal design, and transparency affect users' perception of the chatbot's trustworthiness?*

RQ2: *To what extent do visual design, verbal design, and transparency affect users' experience with the chatbot?*

RQ3: *Do visual design and verbal design interact in their effects on trustworthiness and user experience?*

RQ4: *Do visual and verbal design interact with transparency in their effects on trustworthiness and user experience?*

The results of this study contribute to the academic discussions on human-computer interaction, specifically in two key areas: chatbot design and transparency in service provision. First, it

provides empirical evidence on how different aspects of chatbot design—such as human-like visual and verbal elements—affect user trust and experience. This study adds depth to existing literature by examining not only their combined impact but also how they interact with each other, offering more insights into how users perceive and engage with chatbots.

Second, this research enhances the understanding of transparency in human-machine interactions. By examining how revealing a chatbot's artificial nature affects trust, the study highlights the role of transparency in creating positive service interactions. It also addresses a gap in current research, where the impact of transparency on trust in customer service chatbots has not been studied.

Additionally, this research provides practical guidelines for businesses by highlighting the design and transparency strategies that enhance user trust and experience. The findings can inform the development of more effective chatbot systems, ultimately contributing to improved customer satisfaction and more efficient service delivery in a variety of business settings.

2. Theoretical framework

The next sections establish the study's theoretical framework, starting with defining and examining the dependent variables: trust and user experience. It then focuses on the independent variables—human-like visual and verbal design elements, and transparency—resulting in the development of research hypotheses, questions, and the creation of the research model.

2.1 Trust

Trust is a concept that has been thoroughly examined from multiple angles. One widely recognized view sees trust as a psychological state, where a person is prepared to accept vulnerability, relying on favorable expectations of another's intentions or behavior (Bruckner, 2016; Rousseau et al., 1998). This definition emphasizes the importance of positive expectations and the willingness to be vulnerable, essential components of trust in any context. Comparably, Moorman et al. (1992) describe trust as a dependence on a partner one believes in, highlighting the importance of confidence in a partner's skills and honesty, which is crucial for maintaining effective and lasting partnerships. In the context of chatbots, applying these ideas of trust means that users feel confident in a chatbot's ability to handle tasks and act with their best interests in mind. In addition to that, Mayer et al. (1995, p. 712) define trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party." This definition underscores the readiness to take risks, beyond relying on another's good intentions. As a result, trust in chatbots means users are willing to take the risk of being vulnerable by relying on the chatbot to act in their best interest. This includes ensuring the chatbot is designed to handle sensitive tasks, like securing their data, responsibly.

In human-technology interactions, trust plays a fundamental role in determining the success and adoption of new tools and systems. Trust improves user engagement, satisfaction, and the willingness to rely on chatbot services for various tasks (Song & Shin, 2022; Choudhury & Shamszare, 2023). It influences both the user's intention to adopt the technology and their actual usage of it, impacting overall adoption and effectiveness (Choudhury & Shamszare, 2023). This is crucial because when users feel secure in the technology's ability to protect their data, they are more likely to utilize chatbots for more complex and personal transactions, thereby increasing the utility and integration of these systems in everyday activities.

Besides being seen as a psychological state, trust is often described as a multidimensional construct. These are the attributes that a person needs to have to be perceived as trustworthy. The most widespread one is that characteristics of the trustee comprise three dimensions: Ability, benevolence, and integrity (Mayer et al., 1995). Ability encompasses the skills and competencies necessary to effectively accomplish a specific task. Benevolence involves the trustee's genuine positive intentions that go beyond mere self-interest. Integrity refers to the trustee's adherence to moral principles and fairness, ensuring that their actions are consistent, reliable, and honest (Choung et al., 2022). These three attributes are called the human-like trust constructs (Lankton et al., 2015), and are necessary for one to be able to form trust in relationships.

In this study, human-like trust constructs will be utilized to evaluate the chatbot's trustworthiness. While some researchers have expanded these constructs to include more system-oriented trust models (McKnight et al., 2011), this study focuses on making the chatbot appear as human-like as possible, which is why human-centric trust constructs are more appropriate. In the context of chatbots, ability refers to the chatbot's effectiveness in performing tasks, integrity reflects its

reliability and consistent behavior, and benevolence represents its helpfulness and responsiveness to user needs. Together, these three attributes shape users' perceptions of the chatbot's trustworthiness.

2.2 User experience (UX)

UX refers to the overall way individuals engage with a product, system, or service. UX encompasses both the practical and emotional aspects of interaction, emphasizing the importance of subjective feelings alongside usability and functionality (Hassenzahl, 2003; Law et al., 2009). The International Organization for Standardization (ISO) defines UX as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" (2010, section 2.15), highlighting its broad scope, including pre-use expectations and post-use reflections. Garrett (2011) describes UX as the interplay of various elements, such as information architecture, interface design, and the overall interaction, which collectively shape the user's satisfaction and effectiveness in achieving their goals. Thus, in the context of chatbots, UX encompasses the customer's overall satisfaction and effectiveness in interacting with the chatbot, shaped by both practical usability and emotional responses to the interaction.

The relevance of UX in the context of chatbots is highlighted by its profound impact on multiple important aspects of business-customer relationships. Positive UX has been shown to boost customer satisfaction and foster loyalty, particularly in customer service applications (Jain et al., 2018). Similarly, the growing reliance on chatbots for diverse services underscores the importance of well-designed UX in enhancing efficiency and promoting user acceptance (Følstad & Brandtzæg, 2017). UX not only influences immediate user satisfaction but also plays a critical role in the likelihood of long-term adoption (McLean & Osei-Frimpong, 2019). Key UX elements such as responsiveness, empathy, and contextual understanding are essential for chatbots to meet user expectations and deliver meaningful

value (Hill, Randolph Ford, & Farreras, 2015). Together, these findings establish UX as a cornerstone of effective chatbot design and deployment.

Given the critical role of UX in chatbot adoption and effectiveness, it is essential to assess it when implementing a chatbot. This can be done using the User Experience Questionnaire (UEQ). The UEQ can be used to measure UX across various dimensions, tailored to capture the multifaceted nature of user interactions. Schrepp et al. (2017) describe the UEQ as comprising six key constructs: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. These dimensions collectively provide a complete assessment of both the pragmatic and hedonic parts of user experience.

In the context of chatbots, these constructs are relevant too. Attractiveness evaluates the overall appeal of the chatbot, which can influence initial user acceptance (Følstad & Brandtzæg, 2017). Perspicuity assesses how easily users can understand and learn to use the chatbot, a crucial factor given the diverse user base and varying levels of technological proficiency (Schrepp et al., 2017). Efficiency measures how effectively users can achieve their goals through the chatbot, directly impacting user satisfaction and perceived value (Schrepp et al., 2017). Dependability refers to the chatbot's reliability and predictability, essential for building trust in automated interactions (Schrepp et al., 2017). Stimulation captures the extent to which the chatbot engages and motivates users, enhancing user retention and interaction quality (Schrepp et al., 2017). Lastly, novelty assesses the innovation and creativity of the chatbot, contributing to user interest and differentiation from other services (Schrepp et al., 2017). The UEQ has been proven effective in various fields, showing its strength in evaluating key aspects of user experience, including in the context of chatbots (E te Pas et al., 2020; Laugwitz et al., 2008; Plantak Vukovac et al., 2021). This makes the UEQ a reliable tool for systematically assessing and improving chatbot UX, ensuring they meet user expectations and deliver engaging interactions.

2.3 Humanness in chatbots

A significant concept that has gained prominence in human-machine interaction is anthropomorphism. This refers to the psychological tendency to assign human traits and behaviors to non-human entities (Epley et al., 2007; Seeger et al., 2017). This idea is strongly supported by the media equation theory and the Computers-Are-Social-Actors (CASA) paradigm, which propose that humans naturally respond to robots and machines as though they are interacting with other humans (Kühne & Peter, 2022; Reeves & Nass, 1996). As a result, people often follow social norms and human-like behaviors when engaging with chatbots, perceiving them as social beings (Pitardi & Marriott, 2021). For example, Nass and Moon (2000) found that people tend to be polite to computers and expect similar reciprocal interactions. When users engage with chatbots that exhibit anthropomorphic cues, whether visually or through their language, this leads to expectations of human-like interaction and fosters the perception of humanness (Jackson & Williams, 2021). Overall, anthropomorphism in chatbots has been found to positively influence user perceptions by making interactions feel more natural and intuitive (Araujo, 2018; Melián-González et al., 2019; Nadarzynski et al., 2019; Nass et al., 1994), highlighting its significance in chatbot design.

To fully leverage the benefits of anthropomorphism, chatbots can be crafted to exhibit higher degrees of humanness (Smestad & Volden, 2019). In the realm of chatbots, humanness refers to characteristics that make a chatbot appear more human-like, such as having facial features, eyes, or using polite responses, which are purposefully integrated into its design (Meyer et al., 2016). To effectively replicate human behavior and appearance in chatbots, designers can incorporate different cues. These cues are generally divided into two categories: visual cues and verbal (conversational) cues (Araujo, 2018; Chen et al., 2024; Go & Sundar, 2019). Both types have a crucial part in shaping how people perceive the chatbot and in enhancing its human-like qualities.

In the upcoming sections, these different design cues will be discussed in greater detail, highlighting how they were implemented in this study to maximize the chatbot's perceived humanness, and how they are expected to influence the dependent variables.

2.3.1 Visual design

The appearance of a chatbot is key in influencing users' perception of its human-like characteristics. Visual cues, defined as any observable elements that are visually perceptible aside from words (Chen et al., 2024), are a key factor in making chatbots feel more relatable and engaging. One of the most direct visual cues in chatbot design is its appearance, which can include anything from an abstract shape to a fully anthropomorphic avatar. Research shows that anthropomorphic features, such as human-like avatars, significantly enhance perceptions of humanness in chatbots. For example, studies by Go and Sundar (2019) and Qiu and Benbasat (2009) revealed that chatbots with more human-like appearances enhanced users' perception of social presence, giving them the feeling of interacting with a real person. This increased sense of social presence not only enhances the user experience but also amplifies the chatbot's perceived humanness (Hassanein & Head, 2007). As a result, users tend to respond more positively to chatbots that incorporate anthropomorphic features (Chaves & Gerosa, 2020; Chen et al., 2023; Lee & Oh, 2015; Rhim et al., 2022), as these elements create a more natural and engaging interaction. Furthermore, Parboteeah et al. (2009) found that a visually realistic design can enhance the chatbot's entertainment value, which may positively impact the user experience by increasing its attractiveness, stimulation, and sense of novelty. In sum, designing chatbots with human-like appearances seems to not only improve their perceived humanness and social presence but also align with broader UX goals by enhancing attractiveness, stimulation, and novelty.

Furthermore, the gendered design of a chatbot's visual appearance can influence user perception. Studies suggest that chatbots with a female anthropomorphic design tend to evoke more positive reactions regarding their humanness (Borau et al., 2021). This effect is especially pronounced

for chatbots that offer services, where users may associate female-gendered traits with reliability or warmth (Pawlik, 2022). A more human-like visual design, especially incorporating a gendered avatar, has been shown to foster trust. For example, Chattaraman et al. (2014) demonstrated that a more human-like appearance increased trust in online environments, reinforcing the idea that visual humanness can enhance trustworthiness in chatbot interactions.

Based on these findings, it is expected that using a human-like visual design—specifically a female avatar in this study—can significantly improve users' trust and overall perceptions of the chatbot's user experience. While previous research has generally shown positive effects on both of these outcomes, there is limited understanding of which specific dimensions contribute to these effects and to what degree. This leads to the following hypothesis and related sub-hypotheses:

H1a - H1c: *A more human-like visual design in service chatbots improves the perceived ability, integrity, and benevolence of the chatbot.*

H2a - H2f: *A more human-like visual design in service chatbots improves the perceived attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty of the chatbot.*

2.3.2 Verbal design

The way a chatbot is designed to communicate verbally plays a major role in shaping how users perceive it. Verbal cues, as defined by Feine et al. (2019), refer to social signals conveyed through words. In the case of chatbots, these cues manifest through the conversational style used during interactions. By carefully crafting this aspect of a chatbot's design, designers can significantly enhance its humanness and foster a more positive user experience. For example, the chatbot's conversational style can positively impact customers' enjoyment (hedonic benefits) and practical value (utilitarian benefits)

during the interaction (Wang et al., 2007), thereby enhancing perceptions of attractiveness, stimulation, and novelty—key parts of the user experience.

In order to increase humanness perception in chatbots, designers can focus on implementing "warm" language. Warmth and emotional expressiveness are fundamental human traits that are often perceived as lacking in machines (Borau et al., 2021). While machines are generally assumed to be competent, they tend to fall short in expressing warmth and emotional connection. To address this, employing a warm, friendly conversational style can help balance this perception. Studies indicate that a socially-focused, warm communication style is closely linked to higher trust levels (Bastiansen et al., 2022; Chen et al., 2014; Verhagen et al., 2014). When a chatbot uses friendly and approachable language, users are more inclined to view it as both more human-like and trustworthy (Lee & Choi, 2017), which helps build a stronger connection and enhances the chatbot's reliability during interactions.

Moreover, a chatbot's competence—its ability to efficiently fulfill tasks—can also be reinforced through verbal cues. While users already assume that machines are competent (Borau et al., 2021), reinforcing this through clear, effective communication can enhance perceptions of ability, perspicuity, and efficiency. By effectively handling tasks, the chatbot strengthens users' confidence in its competence, contributing to overall trust in the interaction and the user experience (Roy & Naidoo, 2021). As noted by Kurpicz-Briki (2023), humans often attribute implicit meaning to the chatbot's responses, enhancing the perception of the chatbot's cognitive abilities. Thus, combining warmth with efficient task performance makes sure that the bot is seen as both capable and trustworthy.

Additionally, the use of a personal and friendly tone can impact how users perceive the chatbot's helpfulness and intentions. Friendly communication influences how users view the chatbot's benevolence and integrity—key dimensions of trust (Mayer et al., 1995; Fiske, 2018). If a chatbot seems

personable and caring, users are more likely to believe that it has their best interests in mind, further enhancing trust and dependability.

Incorporating gender cues through language can also increase perceptions of humanness. Historically, warmth has been associated with female personality traits, and a chatbot that introduces itself with a female name, such as “Anna,” may be perceived as friendlier and more human-like (Ellemers, 2018). This not only helps the chatbot seem more approachable but also assigns it an identity, making it easier for users to categorize it as human. In this case, the name serves as both a verbal and identity cue, enhancing the overall sense of humanness (Chen et al., 2024), and making the interaction feel more natural to the user (Ashforth & Humphrey, 1997).

In conclusion, a well-designed verbal approach that combines warm, friendly language with competent task performance, and introduction of the chatbot with a female name, is assumed to boost users’ perceptions of humanness. Additionally, this strategy is likely to strengthen essential trust dimensions—ability, benevolence, and integrity—along with key dimensions of the UX, including attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Consequently, the following hypotheses are proposed:

H3a - H3c: *A more human-like verbal design in service chatbots improves the perceived ability, integrity, and benevolence of the chatbot.*

H4a - H4f: *A more human-like verbal design in service chatbots improves the perceived attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty of the chatbot.*

2.3.3 Congruence between visual and verbal design

Perceptions of chatbot humanness are strongest when a highly human-like visual design is paired with a highly human-like verbal design. According to the theory of avatar marketing, chatbots are most effective when there is alignment between their form realism (visual design) and behavior realism (verbal communication) (Miao et al., 2021). Humans often form expectations about unknown attributes based on known ones (Dick et al., 1990). For instance, a chatbot with a human-like appearance is likely to create an expectation of similarly human-like conversational abilities (Nowak & Biocca, 2003). According to the expectation-confirmation model, when actual experiences align with expectations, it results in positive reinforcement, while a mismatch can have negative consequences (Ambalov, 2018; Bhattacherjee, 2001). Therefore, if customers encounter a highly human-like chatbot avatar paired with equally human-like verbal communication, their expectations are positively confirmed, enhancing their overall experience. Building on the positive effects on user perceptions discussed earlier, this suggests that when both design elements are highly human-like, their combined impact on the dependent variables will be maximized. Thus:

H5a - H5c: *When the level of humanness in visual and verbal design in chatbots align, perceptions of ability, integrity, and benevolence will be improved.*

H6a – H6f: *When the level of humanness in visual and verbal design in chatbots align, perceptions of attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty will be improved.*

2.4 Transparency

Transparency, when it comes to chatbots, means how openly it is conveyed to users that they are engaging with an automated system rather than a real person. In chatbot interactions, transparency can help manage expectations and foster trust, particularly when the system openly discloses its

automated nature. For instance, research has discovered that when users are told they are engaging with a machine, it mitigates the risk of users feeling deceived, thereby strengthening trust in the system's integrity and benevolence (Bejger & Elster, 2020). Additionally, transparency can enhance perceived ability, as users who are aware of the chatbot's limitations and strengths are more likely to trust its competence (Hoff & Bashir, 2015). Moreover, transparency aligns with broader principles of trustworthy AI systems, which emphasize the importance of clearly communicating system capabilities and constraints to users. By fostering an open interaction environment, transparency reduces uncertainty and enables users to form more accurate trust judgments (Lee & See, 2004). This promotes appropriate trust, where users neither over-rely nor under-rely on the chatbot based on realistic expectations of its abilities (Crockett et al., 2020), which helps reduce frustrations during interactions as well (Chen, 2022; Khurana et al., 2021).

Building on these findings, it is hypothesized that disclosing the chatbot's artificial nature to users will result in higher perceived trust, specifically in terms of its ability, benevolence, and integrity. Since the second dependent variable in this study is user experience (UX), it is especially interesting to explore whether transparency affects different aspects of UX. This has not been studied much in scientific research, so it could offer new and useful insights. Furthermore, human-like design elements have the potential to make the chatbot seem more like a real person, which raises the question whether this conflicts with or complements the trust gained from the transparency about its artificial nature. The existing literature has yet to thoroughly explore this potential interaction between transparency and human-like design elements. Thus, the following hypotheses and research questions are proposed:

H7a – H7c: *Disclosing the chatbot's artificial nature improves the perceived ability, integrity, and benevolence of the chatbot.*

SQ1a-SQ1f: *Does the disclosure of the chatbot's artificial nature effect the perceptions of attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty?*

SQ2: *Does the disclosure of the chatbot's artificial nature influence the effects of human-like visual design of the chatbot on the perceived trust and user experience dimensions?*

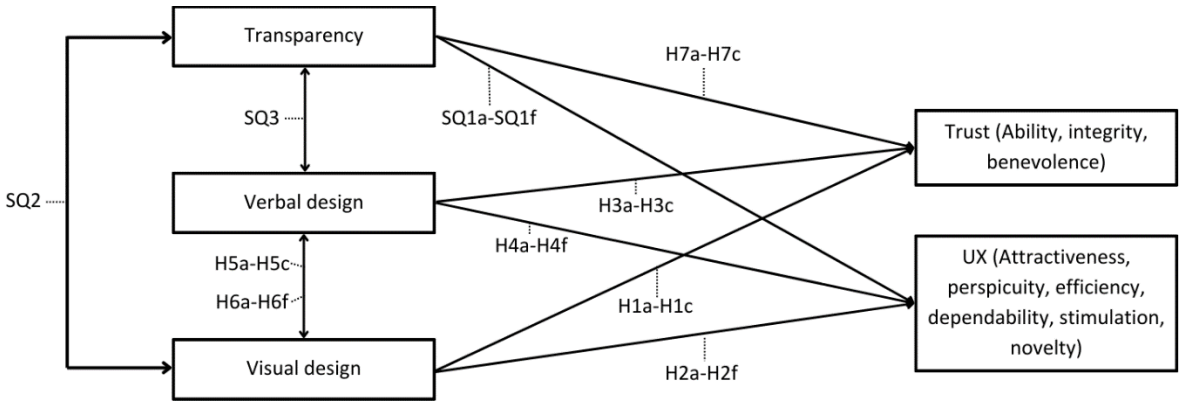
SQ3: *Does the disclosure of the chatbot's artificial nature influence the effects of human-like verbal design of the chatbot on the perceived trust and user experience dimensions?*

2.5 Conclusion

The theoretical framework explored the concept of trust and user experience, and their relevance in chatbot interactions. Furthermore, visual design, verbal design, and transparency were introduced as possible predictors of trust and UX. Based on this the hypotheses were proposed, forming the research design, as illustrated in Figure 1.

Figure 1

Research model



3. Methodology

The study's methodology began with a qualitative pre-test to identify the most effective human-like design elements for the study's context. Insights from this pre-test informed the design of the main study, which was conducted using a quantitative approach. The following sections will first provide a detailed description of the pre-test, followed by an explanation of the methodology used in the primary study.

3.1 Pre-test

The pre-test was conducted to determine which visual and verbal adjustments most effectively influenced perceptions of human-like qualities and robotic characteristics in the chatbot's design. This step was crucial for refining the manipulations before the main study. A qualitative approach was chosen to gather detailed insights and feedback from participants. Qualitative research focuses on understanding how people experience social phenomena through conversations or observations (Malterud, 2001) and provides insights into complex contexts that quantitative methods may overlook (Doz, 2011). This approach allowed the pre-test to explore participants' perceptions, leading to better-informed design adjustments.

3.1.1 Pre-test materials

Four distinct characters were designed to represent varying degrees of human-likeness for the visual design of the chatbot. These characters were created using *Google*'s "Imagen Generator," an AI-based image generation tool. Figure 2 illustrates the different avatars, which range in appearance from highly cartoonish to highly human-like.

Figure 2

Chatbot avatars for pre-testing

Note. The avatars are arranged from left to right, with the leftmost being the most cartoonish and the rightmost being the most human-like.

Participants were shown each character within the chatbot design (see Figure 3 for an example). Then they were asked to say what they thought of each character, focusing on their human-like qualities and how well they fit within the chatbot context. The aim was to gather qualitative insights into which visual design most strongly conveyed this attribute and felt the most natural to the participants.

Figure 3

Example of an avatar in the chatbot design

Note. This example shows the second avatar, which is the slightly less cartoonish version.

In addition to the visual stimuli, participants were introduced to the first version of the verbal chatbot design. Two different scripts for the chatbot dialogue were created with the help of *OpenAI*'s *ChatGPT 4o* to test the verbal design: one designed to be human-like and the other designed to be robotic. The human-like script employed a "warm" language, incorporating conversational elements and empathetic responses, which made it seem friendly and personal. In contrast, the robotic script utilized straightforward, "cold" language, minimizing social cues and emotional engagement.

The human-like script was designed to mimic the warmth and engagement of human conversation. First, personalized language was utilized, incorporating personal pronouns whenever

possible. For example, phrases like “I’m here to help you” and “Can you confirm if you’re experiencing this issue on all your devices?” create a sense of individualized attention and increase the dialogue quality (Caldarini et al., 2022). Secondly, the chatbot expressed understanding and empathy, which is crucial for building a connection with users (Aldrup et al., 2022). Empathy in chatbot design helps in creating a supportive interaction, fostering a sense of emotional warmth (He et al., 2023). This was evident in responses such as “I’m sorry to hear that. Let’s see what we can do to improve your connection,” demonstrating a concern for the user’s issues. Finally, warm language is characterized by high levels of politeness (Jeong et al., 2019), aiming to make users feel valued and respected. This was reflected in statements such as “That’s wonderful to hear! If you encounter any more issues, feel free to reach out.”

The robotic script was designed to be “cold” and devoid of unnecessary social niceties. To achieve the robotic design, the chatbot’s dialogue was crafted with several key characteristics that are the opposite of the ones creating the “warm” language. The chatbot used direct commands, employing imperative statements without softening the language. For instance, phrases like “Provide details of your issue” and “Restart your router and check your connection” were used to convey clear instructions. Additionally, the script intentionally avoided any expressions of empathy or emotion, creating a more detached interaction style. Instead of saying “I’m sorry to hear that,” the chatbot would respond with “Acknowledged,” thus maintaining a neutral and unemotional tone. Responses were concise and directed solely at addressing the task at hand. An example of this is seen in statements like “Commencing reset. Wait a minute.” Finally, the language was formal and impersonal, deliberately avoiding elements that would imply a personal touch. For example, instead of saying “That’s wonderful to hear!” the chatbot would use “Confirmed. Connection improved,” emphasizing a functional rather than relational interaction.

3.1.2 Pre-test procedure

Participants took part in a semi-structured interview lasting about 15 minutes. Ten participants were selected for the pre-test, chosen based on their availability and interest in participating in qualitative research. Their ages ranged from 22 to 63, and their nationalities were predominantly Dutch and German. During the interview, the researcher took detailed notes to capture key points for later analysis. The data was then analyzed by the researcher, ensuring a consistent interpretation aligned with the study's objectives.

At the start of the interview, participants were informed about the intent of the research and asked for verbal consent (see Appendix A). Initially, they were presented with the four different visual designs of the chatbot in random order. Participants were then asked to share their perceptions of these designs, identifying which characters they perceived as the most human-like and which they thought fit best in the chatbot context.

Following the visual assessment, participants were shown the two versions of the chatbot script. This was done in text form to ensure that their perception of the script was not influenced by the chatbot's visual design, whether robotic or human-like. Participants were asked to comment on the perceived human-likeness or roboticness of the chatbot's responses and to evaluate the effectiveness of the language used in conveying these traits. They also provided suggestions for improving the chatbot's verbal interactions to better align with the desired perceptions.

The questions prepared for the semi-structured interview can be found in Appendix A. Some examples are "Considering the chatbot context, which avatar do you think fits best as a chatbot? What makes you feel this way?", and "What suggestions do you have for improving the chatbot's verbal interactions to better convey the desired human-like or robotic traits?" These questions guided the

participants through their evaluation of both the visual and verbal designs, ensuring comprehensive feedback on the chatbot's human-like and robotic traits.

3.1.3 Pre-test results

The pre-test yielded several findings regarding the participants' perceptions of the chatbot's visual and verbal manipulations. Participants were able to consistently identify which characters seemed the least human and which seemed the most human. Notably, seven out of the ten participants preferred the second avatar, which was the slightly less cartoonish version, for the context of a human-like service chatbot. They felt that this avatar struck a balance between being relatable and maintaining a professional appearance. The more cartoonish avatar was seen as less suitable for a service context, while the more human-like avatars were perceived as too realistic, potentially causing uncanny valley effects.

All participants could clearly distinguish between the two scripts. They described the human-like script as seemingly more "human-written," feeling more personal and engaging, whereas the robotic script was characterized as very straightforward and cold sounding. Participants appreciated the conversational tone and empathetic responses of the human-like script, noting that it made the interaction feel more personalized and supportive. One critique mentioned was that the interaction would feel more natural if the chatbot introduced itself by name at the beginning of the conversation.

While participants agreed that the straightforward nature of the robotic script was effective, some suggested that the responses could be made even shorter to seem even more cold. For example, instead of saying "Commencing reset. Wait a minute," it could be condensed to "Resetting. Wait." Overall, participants found the scripts worked well in conveying the desired human-like and robotic traits.

Based on this feedback, the second leftmost avatar, which is slightly less cartoonish, was chosen for the primary study. Additionally, the robotic script was refined to be even more straightforward and to the point, while the human-like script remained mostly unchanged, except for one slight adjustment: the bot now introduces itself by name at the beginning of the conversation. The final version of the scripts can be found in Appendix A.

3.2 Primary study

3.2.1 Design

This quantitative study utilized a 2x2x2 factorial experimental design. The independent variables included the chatbot's visual design (highly human-like vs. less human-like), its verbal design (highly human-like vs. less human-like), and transparency (transparent vs. not transparent). The study's goal was to examine how these design variations influenced the dependent variables: perceived trustworthiness and user experience. Additionally, it investigated possible interaction effects between visual design, verbal design, and transparency on these outcomes. A summary of all experimental conditions is provided in Table 1.

Table 1

Experimental conditions

Experimental condition	Visual design	Verbal design	Transparency
1	Highly human-like	Highly human-like	Not transparent
2	Highly human-like	Highly human-like	Transparent
3	Highly human-like	Less human-like	Not transparent
4	Highly human-like	Less human-like	Transparent
5	Less human-like	Highly human-like	Not transparent
6	Less human-like	Highly human-like	Transparent
7	Less human-like	Less human-like	Not transparent
8	Less human-like	Less human-like	Transparent

3.2.2 Instruments

Participants watched a video showcasing a chatbot interaction, where both the visual and verbal designs had been adjusted. The videos were created using *Canva* premium, a graphic design platform for the creation of content (https://www.canva.com/). The visual design includes a human-like name and avatar for the chatbot, while in the less human-like condition, a blank spot replaces the avatar, and the chatbot is named “chat assistant.” Verbal design varies with warm and friendly language in human-like conditions and robotic, cold language in the less human-like condition. Examples of these variations are shown in Figure 4.

Figure 4

Examples of experimental condition 1 and 7

Note. The phone on the left displays a picture from the video of condition one, featuring a highly human-like verbal and visual design. The phone on the right shows condition seven, with a less human-like verbal and visual design.

Transparency about the chatbot interaction is a central focus of the study. Every participant, regardless of the condition they are assigned to, receives an identical informed consent form, which does not explicitly state that their interaction will be with a chatbot. Participants in the non-transparent conditions are not informed at any point before they watch the video that they will see an interaction with a chatbot; the informed consent form and the brief message before the video only mention sharing thoughts on an online service experience, omitting the detail about the chatbot interaction.

In contrast, participants in the transparent conditions are informed before watching the video that they will watch an interaction with a chatbot. This guarantees that participants are fully informed about the nature of the interaction from the very beginning. After completing the entire survey, participants in both conditions are debriefed about the true nature of their interaction and the reason for the initial omission in the non-transparent conditions. This approach is employed because the study poses no potential risks or harm to the participants, and their unawareness is crucial for maintaining the integrity of the research findings. The distinction between the groups allows for the examination of how transparency about interacting with a chatbot versus the absence of such information influences participants' perceptions.

3.2.3 Measures

Trust in the chatbot was evaluated using a modified version of the scale developed by Schoorman et al. (1996), which measures key factors like ability, benevolence, and integrity—traits known to influence trust in technology. Participants responded to statements on a five-point Likert scale, ranging from "Strongly disagree" to "Strongly agree." Sample statements included, "The chatbot is very capable of performing its tasks" and "The chatbot goes out of its way to help me." The full measurement scale is provided in Appendix B, Table B2.

User experience was measured with the User Experience Questionnaire by Laugwitz et al. (2008), which evaluates dimensions such as attractiveness, perspicuity, novelty, stimulation, dependability, and efficiency. Participants rated their experiences using a seven-point semantic differential scale, selecting between pairs of opposite adjectives in relation to the chatbot or chat assistant (in non-transparent conditions). Examples include "unattractive - attractive" and "boring - exciting." All the items used for measurement are detailed in Appendix B, Table B3.

3.2.4 Participants

Participants for the study were recruited through online platforms and social media channels, with adherence to ethical guidelines requiring all participants to be at least 18 years of age. While demographic variables such as gender, educational level, and nationality are not directly relevant to the established research design, they were recorded for the possibility of additional analyses should interesting patterns emerge (see Appendix B, Table B1 for the detailed questions).

Out of the total 260 responses collected, 68 had to be excluded from the final dataset due to missing values or invalid entries, such as failure to accept the informed consent or to finish the entire survey. This left the final dataset with 192 responses. The sample consisted of individuals with ages ranging from 18 to 66, and a mean age of 27.55 years (SD = 7.56). The distribution of participants by nationality was as follows: 23.96% German (46 participants), 6.25% Dutch (12 participants), and 69.79% from other nationalities (134 participants). Regarding the highest level of education, the sample included 11.46% with a high school diploma (22 participants), 46.35% with a Bachelor's degree (89 participants), 35.42% with a Master's degree (68 participants), 4.69% with a Doctorate (9 participants), and 2.08% with other educational qualifications (4 participants). Table 2 shows an overview of the mean age and the nationalities and educational levels per condition.

Table 2

Overview of demographics per condition

Condition	Mean age	Nationality	Educational level
Human-like visual design, human-like verbal design & no transparency (n = 21)	28.0	Other (n = 16) German (n = 5) Dutch (n= 0)	High school (n = 2) Bachelor's degree (n = 9) Master's degree (n = 7) Doctorate (n = 3) Other (n = 0)
Human-like visual design, human-like verbal design & transparency (n = 27)	27.9	Other (n = 22) German (n = 4) Dutch (n = 1)	High school (n = 3) Bachelor's degree (n = 13) Master's degree (n = 8) Doctorate (n = 3) Other (n = 0)
Human-like visual design, not human-like verbal design & no transparency (n = 27)	26.0	Other (n = 16) German (n = 6) Dutch (n = 5)	High school (n = 2) Bachelor's degree (n = 12) Master's degree (n = 10) Doctorate (n = 1) Other (n = 2)
Human-like visual design, not human-like verbal design & transparency (n = 26)	27.0	Other (n = 18) German (n = 6) Dutch (n = 2)	High school (n = 3) Bachelor's degree (n = 10) Master's degree (n = 11) Doctorate (n = 2) Other (n = 0)
Not human-like visual design, human-like verbal design & no transparency (n = 24)	27.6	Other (n = 16) German (n = 8) Dutch (n = 0)	High school (n = 3) Bachelor's degree (n = 13) Master's degree (n = 8) Doctorate (n = 0) Other (n = 0)
Not human-like visual design, human-like verbal design & transparency (n = 21)	27.9	Other (n = 17) German (n = 3) Dutch (n = 1)	High school (n = 3) Bachelor's degree (n = 13) Master's degree (n = 5) Doctorate (n = 0) Other (n = 0)
Not human-like visual design, not human-like verbal design & no transparency (n = 24)	29.6	Other (n = 15) German (n = 8) Dutch (n = 1)	High school (n = 2) Bachelor's degree (n = 10) Master's degree (n = 11) Doctorate (n = 1) Other (n = 0)
Not human-like visual design, not human-like verbal design & transparency (n = 22)	26.7	Other (n = 15) German (n = 5) Dutch (n = 2)	High school (n = 4) Bachelor's degree (n = 9) Master's degree (n = 8) Doctorate (n = 1) Other (n = 0)

3.2.5 Procedure

Participants received a survey link and were randomly assigned to one of the eight conditions, corresponding to the 2x2x2 factorial design. After giving informed consent, they answered some general demographic questions (Appendix B, Table 1B). Following this, they viewed a pre-recorded video of a chatbot interaction that matched their assigned condition. To provide participants with a detailed context for the survey, two distinct scenarios were crafted based on whether they were to be aware of interacting with a chatbot or not. For participants unaware they were watching an interaction with a chatbot the scenario was "Imagine you're experiencing issues with your internet service provided by a specific company, and you're keen on resolving these issues promptly. To seek assistance, you turn to the company's website and decide to use their chat option, hoping for a quick and efficient resolution. On the next page, you will watch a video depicting this service interaction. Please observe the interaction closely, as you will be asked to share your thoughts and impressions afterwards. [...]."

For participants aware they were watching an interaction with a chatbot it was: "Imagine you're experiencing difficulties with your internet service from a certain company and are eager to find a solution. You choose to engage with the company's chat service for help, fully aware that this service is a chatbot designed to assist customers. On the next page, you will watch a video depicting this chatbot interaction. Please observe the interaction closely, as you will be asked to share your thoughts and impressions afterwards. [...]."

After watching the video, participants filled out a survey with scales for perceived trustworthiness and user experience. All in all, the study took around 10 minutes to complete.

3.2.6 Data analysis

The dataset was refined by removing extraneous information, including response type, start and end dates, location latitude and longitude, recorded date, and so on. All statistical analyses were conducted exclusively on this cleaned dataset, and significance was set at $p < .05$.

Statistical regression analysis was used to evaluate the main effects of visual design, verbal design, and transparency on perceived trustworthiness and user experience. Interaction effects were also analyzed using analyses of variance to determine whether combining design manipulations and transparency had combined effects on the dependent variables. The analysis was conducted using *R* and *RStudio*.

The measurement scales for perceived trust and user experience were adapted from established scales in human-computer interaction and service research to ensure content validity. Additionally, a reliability analysis was performed to evaluate the internal consistency of the items assessing three dimensions of trust—ability, benevolence, and integrity—and six dimensions of user experience: attractiveness, perspicuity, novelty, stimulation, dependability, and efficiency. The Cronbach's alpha values for the individual scales are presented in Table 3.

Table 3

Cronbach's alpha results

Scales	**Cronbach's Alpha (α)**
Ability (Six items)	.85
Benevolence (Five items)	.73
Integrity (Six items)	.83
Attractiveness (Six items)	.90
Perspicuity (Four items)	.85
Novelty (Four items)	.82
Stimulation (Four items)	.83
Dependability (Four items)	.76
Effectiveness (Four items)	.74

These results suggest that the items within each scale reliably measure their respective constructs, making them suitable for further analysis in the context of trust and user experience measurement.

4. Results

4.1 Human-like visual design

The hypotheses proposed that a more human-like verbal design in service chatbots leads to higher perceived ability (H1a), integrity (H1b), and benevolence (H1c). Table 4 depicts a summary of the regression analysis results for these hypotheses.

Table 4

Regression analysis results for human-like visual design and trust dimensions

Hypothesis	*F*	df1	df2
H1a. *A more human-like visual design in service chatbots improves the perceived ability of the chatbot.*	< 1	1	190
H1b. *A more human-like visual design in service chatbots improves the perceived integrity of the chatbot.*	< 1	1	190
H1c. *A more human-like visual design in service chatbots improves the perceived benevolence of the chatbot.*	< 1	1	190

In summary, the results do not support any of the hypotheses. A more human-like visual design in service chatbots does not significantly enhance trust in terms of perceived ability, integrity, or benevolence.

The next hypotheses propose that a more human-like visual design in service chatbots leads to higher perceived user experience, specifically in terms of attractiveness (H2a), perspicuity (H2b), efficiency (H2c), dependability (H2d), stimulation (H2e), and novelty (H2f). None of the hypotheses were supported, as human-like visual design failed to significantly predict any of the individual dimensions. These results are summarized in Table 5.

Table 5

Regression analysis results for human-like visual design and UX dimensions

Hypothesis	*F*	df1	df2	*p*
H2a. *A more human-like visual design in service chatbots improves the perceived attractiveness of the chatbot.*	<1	1	190	.82
H2b. *A more human-like visual design in service chatbots improves the perceived perspicuity of the chatbot.*	< 1	1	190	.79
H2c. *A more human-like visual design in service chatbots improves the perceived efficiency of the chatbot.*	1.64	1	190	.20
H2d. *A more human-like visual design in service chatbots improves the perceived dependability of the chatbot.*	< 1	1	190	.97
H2e. *A more human-like visual design in service chatbots improves the perceived stimulation of the chatbot.*	< 1	1	190	.35
H2f. *A more human-like visual design in service chatbots improves the perceived novelty of the chatbot.*	1.04	1	190	.31

To sum up, these results do not support any of the hypotheses. A more human-like visual design in service chatbots does not significantly enhance user experience in terms of attractiveness, perspicuity, efficiency, dependability, stimulation and novelty.

4.2 Human-like verbal design

The hypotheses proposed that a more human-like verbal design in service chatbots leads to higher perceived trust, specifically through higher perceived ability (H3a), integrity (H3b), and benevolence (H3c). The results of the regression analyses conducted to test these hypotheses can be found in Table 6.

Table 6

Regression analysis results for human-like verbal design and trust dimensions

Hypothesis	*F*	df1	df2	*p*
H3a. *A more human-like verbal design in service chatbots improves the perceived ability of the chatbot.*	1.43	1	190	.23
H3b. *A more human-like verbal design in service chatbots improves the perceived integrity of the chatbot.*	1.33	1	190	.25
H3c. *A more human-like verbal design in service chatbots improves the perceived benevolence of the chatbot.*	5.86	1	190	.02

In conclusion, while a more human-like verbal design in service chatbots did not increase perceptions of ability or integrity, it significantly enhanced perceived benevolence, supporting hypothesis H3c.

The next hypotheses propose that a more human-like verbal design in service chatbots leads to higher perceived user experience, in terms of attractiveness (H4a), perspicuity (H4b), efficiency (H4c), dependability (H4d), stimulation (H4e), and novelty (H4f). Table 7 provides a summary of the regression analysis results for these hypotheses.

Table 7

Regression analysis results for human-like verbal design and UX dimensions

Hypothesis	*F*	df1	df2	*p*
H4a. *A more human-like verbal design in service chatbots improves the perceived attractiveness of the chatbot.*	29.17	1	190	< .001
H4b. *A more human-like verbal design in service chatbots improves the perceived perspicuity of the chatbot.*	2.12	1	190	.15
H4c. *A more human-like verbal design in service chatbots improves the perceived efficiency of the chatbot.*	2.38	1	190	.13
H4d. *A more human-like verbal design in service chatbots improves the perceived dependability of the chatbot.*	4.29	1	190	.04
H4e. *A more human-like verbal design in service chatbots improves the perceived stimulation of the chatbot.*	13.11	1	190	< .001
H4f. *A more human-like verbal design in service chatbots improves the perceived novelty of the chatbot.*	11.55	1	190	< .001

In summary, the results partially support the hypotheses. A more human-like verbal design significantly improved attractiveness, dependability, stimulation, and novelty but did not significantly affect perspicuity or efficiency.

In Table 8 an overview of the mean scores and standard variations for the variables that were significantly affected by the human-like verbal design manipulations can be found.

Table 8

Mean scores and standard deviations for significant verbal design manipulation effects

Variable	Not human-like verbal design		Human-like verbal design	
	M	***SD***	***M***	***SD***
Benevolence	3.42	0.73	3.67	0.72
Attractiveness	4.50	1.22	5.32	0.87
Dependability	5.22	0.99	5.50	0.85
Stimulation	4.23	1.18	4.80	1.01
Novelty	4.00	1.21	4.57	1.14

4.3 Transparency

The hypotheses suggested that disclosing a chatbot's artificial nature leads to higher perceived ability (H7a), integrity (H7b), and benevolence (H7c). A summary of the regression analyses results for these hypotheses is illustrated in Table 9.

Table 9

Regression analysis results for transparency and trust dimensions

Hypothesis	***F***	df1	df2	***p***
H7a. *Disclosing the chatbot's artificial nature improves the perceived ability the chatbot.*	4.48	1	190	.04
H7b. *Disclosing the chatbot's artificial nature improves the perceived integrity of the chatbot.*	6.44	1	190	.01
H7c. *Disclosing the chatbot's artificial nature improves the perceived benevolence of the chatbot.*	<1	1	190	.44

In conclusion, the results partially support the hypotheses. While transparency about a chatbot's artificial nature significantly improves trustworthiness in terms of perceived ability and integrity, it does not significantly affect perceived benevolence.

The next sub-questions explored whether being transparent about the chatbots artificial nature leads to higher perceived user experience, in terms of attractiveness (SQ1a), perspicuity (SQ1b), efficiency (SQ1c), dependability (SQ1d), stimulation (SQ1e), and novelty (SQ1f). Table 10 provides a summary of the regression analysis results for these questions.

Table 10

Regression analysis results for transparency and UX dimensions

Hypothesis	***F***	**df1**	**df2**	***p***
SQ1a. *Does the disclosure of the chatbot's artificial nature effect the perceptions of attractiveness?*	1.17	1	190	.28
SQ1b. *Does the disclosure of the chatbot's artificial nature effect the perceptions of perspicuity?*	2.43	1	190	.12
SQ1c. *Does the disclosure of the chatbot's artificial nature effect the perceptions of efficiency?*	1.16	1	190	.28
SQ1d. *Does the disclosure of the chatbot's artificial nature effect the perceptions of dependability?*	2.16	1	190	.14
SQ1e. *Does the disclosure of the chatbot's artificial nature effect the perceptions of stimulation?*	2.07	1	190	.15
SQ1f. *Does the disclosure of the chatbot's artificial nature effect the perceptions of novelty?*	2.97	1	190	.09

To summarize, these results show that none of the sub-questions 1 have a positive answer. Being transparent about the chatbot's nature does not significantly affect user experience in terms of attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty.

Table 11 summarizes the mean scores and standard deviations of the variables that showed significant effects from the transparency manipulations.

Table 11

Mean scores and standard deviations for significant transparency manipulation effects

Variable	No transparency		Transparency	
	M	***SD***	***M***	***SD***
Ability	3.71	0.79	3.93	0.63
Integrity	3.56	0.78	3.82	0.64

4.4 Interaction effects

4.4.1 Interaction between human-like visual and verbal design

A two-way analysis of variance (ANOVA) was used to explore how the interaction between visual design (human-like vs. non-human-like) and verbal design (human-like vs. non-human-like) influenced the dependent variables. The analysis aimed to test hypotheses H5a to H5c, and H6a to H6f, which proposed that perceptions of trust and user experience dimensions would be higher when the level of humanness in visual and verbal design aligns.

The first two-way ANOVAs revealed no significant interaction effects between visual and verbal design on any of the trust dimensions, with all F-values being $F(1, 188) < 1$, indicating that the combined effect of visual and verbal design on ability, integrity and benevolence does not depend on the alignment of these two factors.

Similarly, the second two-way ANOVAS for the user experience dimensions did not yield significant interaction effects between visual and verbal design, with all F-values again being $F(1, 188) <$

1. This indicates that the alignment of visual and verbal humanness does not significantly influence any of the user experience aspects. Based on this, all the H5 and H6 hypotheses were rejected.

4.4.2 Interaction between transparency and human-like visual design

The two-way ANOVAs did not reveal any significant interaction effects between transparency and human-like visual design on any of the trust dimensions. The interaction effect on perceived ability had an F-value of $F(1, 188)= 1.24$, $p = .27$, integrity was $F(1, 188) < 1$, and benevolence $F(1, 188) = 1.59$, $p = .21$. Similarly, the interaction effects on the user experience dimensions were not significant. The F-values and p-values can be seen in Table 12.

Table 12

ANOVA results for interaction effects between transparency and visual design on UX dimensions

UX dimension	***F***	**df1**	**df2**	***p***
Attractiveness	1.44	1	188	.23
Perspicuity	1.16	1	188	.28
Efficiency	< 1	1	188	.45
Dependability	< 1	1	188	.54
Stimulation	1.17	1	188	.28
Novelty	< 1	1	188	.26

This analysis addressed SQ2, which explored whether revealing that the chatbot is artificial affects how its human-like visual design influences any of the trust and user experience dimensions. The results show that this disclosure has no impact.

4.4.3 Interaction between transparency and human-like verbal design

The two-way ANOVA did not reveal any significant interaction effects between transparency and human-like verbal design on the trust dimensions. The interaction effect on perceived ability had an F-

value of F(1, 188)= 1.23, *p* = .27, integrity was F(1, 188) = 1.68, *p* = .20, and benevolence F(1, 188) = < 1. The interaction effects on any of the user experience dimensions were also not significant, an overview of the F and p-values per dimension can be found in Table 13.

Table 13

ANOVA results for interaction effects between transparency and verbal design on UX dimensions

UX dimension	***F***	**df1**	**df2**	***p***
Attractiveness	1.40	1	188	.24
Perspicuity	1.81	1	188	.18
Efficiency	1.97	1	188	.16
Dependability	1.48	1	188	.23
Stimulation	< 1	1	188	.82
Novelty	1.29	1	188	.26

Here, question SQ3 was addressed. SQ3 explored whether revealing the chatbot's artificial nature influenced the impact of its human-like verbal design on any of the trust and user experience dimensions. This question was also answered negatively.

5. Discussion

This study investigated how human-like visual and verbal design, as well as transparency about a chatbot's artificial nature, influence perceived trust and user experience in service chatbots. Key findings indicate that human-like visual design did not significantly enhance trust or user experience. In contrast, human-like verbal design positively impacted user experience—particularly in attractiveness, dependability, stimulation, and novelty—and improved perceived benevolence. Transparency about the chatbot's artificial nature increased trust in terms of ability and integrity but had no effect on benevolence or user experience dimensions. Furthermore, no interaction effects were observed between visual and verbal design or between transparency and design cues, suggesting these elements operate independently in shaping trust and user experience.

5.1 Human-like design elements

Contrary to the expectations derived from the theoretical framework, the findings show that human-like visual design in service chatbots did not significantly impact any of the trust or UX dimensions. However, the results showed that the verbal design of the chatbot did have effects on some of the dimensions. This leads to the assumption that trust and user experience might depend more on the chatbot's conversational abilities than its visual design, which has been suggested by other researchers before (Blascovich et al., 2002).

In addition to that, research has shown that the effectiveness of anthropomorphic design (including visual human-likeness) could depend on task relevance. Gong (2008) found that anthropomorphism is more effective in social or emotional contexts, where the human-like qualities help establish rapport, rather than in functional or transactional contexts. This aligns with cognitive load theory (Sweller, 1988), which suggests that users might prioritize clarity and simplicity when interacting with technology, especially in task-based contexts. If the chatbot's primary role is to assist with tasks,

solve problems, or provide information, which is the case in this study, a human-like appearance might not align with user expectations, thereby failing to enhance user experience. This seems to be similar for the effects of the verbal design, since the more hedonic aspects of UX were improved when a more human-like conversational style was used, while the more utilitarian ones—perspicuity and efficiency—were not affected.

As specified by media richness theory (Daft & Lengel, 1986), communication mediums that convey social cues and emotional undertones are better suited for complex or ambiguous tasks. Human-like verbal design adds social richness to chatbot interactions, enhancing user experience by making the chatbot feel more engaging and relatable, particularly in areas like attractiveness and stimulation. This may explain why users rated chatbots with human-like verbal design higher in terms of these dimensions. These aspects of UX are closely tied to the emotional experience of interacting with the chatbot. In this study, these emotional cues were conveyed primarily through conversation rather than visual design, which could explain why only the verbal design had this impact.

Looking more at trust, these findings also align with theories that distinguish between emotional (benevolence) and cognitive (ability and integrity) components of trust (Mayer et al., 1995). Verbal design, particularly when human-like, may be more effective at conveying warmth, friendliness, and empathy—qualities closely associated with benevolence—than at signaling competence or integrity. This is consistent with research showing that anthropomorphized agents can evoke a sense of emotional connection and perceived care, which enhances benevolence (Waytz, Heafner, & Epley, 2014). Participants perceived the chatbot with a more human-like conversational style as more benevolent in the task-focused context. However, for ability and integrity—traits more reliant on performance—users may prioritize the chatbot's clear problem-solving skills and accuracy over the verbal tone or human-like phrasing.

Additionally, users may be familiar with engaging with non-human-like digital interfaces, such as text-based chatbots or simple virtual assistants like Apple's Siri or Amazon's Alexa. Studies indicate that familiarity with simpler interfaces may lead to more positive interactions, as users do not expect or need high levels of anthropomorphism (Nass & Moon, 2000). If users have learned to expect functionality over human-like appearances in these interactions, the human-like visual design could be irrelevant or even distracting. While many factors could play a role in why a more human-like visual design in chatbots did not lead to higher trust and user experience, it becomes clear that the visual design of the bot seems to play a secondary role in the design and effectiveness of a chatbot.

Moreover, users seem to evaluate visual and verbal elements independently rather than as a cohesive whole, which implies that efforts to synchronize these elements may not yield additional benefits in terms of enhancing trust or user experience. Although this finding was initially unexpected, given previous research suggesting that aligning an avatar's visual realism with its behavioral design enhances its effectiveness (Miao et al., 2021), the dual-coding theory provides insight into why no such effect was observed in this case. It posits that humans process visual and verbal information through distinct cognitive systems, but it also suggests that these systems can work together to enhance understanding and memory when the information is complementary (Paivio, 1986). However, in this study's context, the visual and verbal elements may not have been perceived as complementary or relevant to each other since they were more focused on the interaction itself. This may have led users to evaluate them independently. This could explain why aligning human-like features across both dimensions did not significantly enhance trust or user experience. Depending on the context of chatbot use (e.g., customer service), users may have developed specific expectations where they place more emphasis on verbal communication than on visual aspects (Gong, 2008). In such cases, users might expect chatbots to be clear and helpful rather than visually realistic, meaning that the alignment between verbal and visual design would not necessarily boost trust or user experience.

5.2 Transparency

The results show that being transparent about the chatbot's artificial nature enhances trust, specifically by improving perceived ability and integrity, but not benevolence. Trust in automation theory (Lee & See, 2004) emphasizes that trust in technology is often formed based on predictability and reliability. According to the expectation-confirmation model, people set expectations about unknown factors based on something they know (Ambalov, 2018; Bhattacherjee, 2001). Transparency about a chatbot's artificial nature may lead users to set realistic expectations about its abilities, particularly in terms of competence. When users know the chatbot is artificial, they can predict its performance more accurately, often seeing it as efficient and capable within its limits. Transparency can also enhance perceptions of integrity, as users may view the chatbot's honesty about its nature as ethical and reassuring (Xu et al., 2023). This aligns with findings that transparency reduces cognitive dissonance, as users are less likely to feel misled about the chatbot's capabilities, improving their perception of its honesty (Madhavan & Wiegmann, 2007). Transparency might also enhance perceived integrity because it aligns with the principle of ethical AI (Binns et al., 2018), where disclosing the artificial nature of technology is seen as promoting honesty and openness. This transparency could lead users to view the chatbot as more trustworthy in terms of uprightness, interpreting the disclosure as a signal of ethical practice, which enhances perceptions of integrity. Moreover, these perceptions might even be extended towards the company behind the bot.

Transparency about the chatbot's artificial nature did not enhance any UX dimensions, which can be explained through the customer journey perspective (Lemon & Verhoef, 2016). According to this perspective, factors like transparency is likely more impactful during pre- or post-interaction phases, where it helps set expectations and reduce cognitive dissonance (Lemon & Verhoef, 2016)—effects that align with the observed improvements in trust. However, during the interaction itself, transparency

becomes less critical. Since UX dimensions are inherently tied to the immediate interaction experience, transparency may not directly influence these aspects, explaining its limited effect in this context.

The lack of interaction effects suggests that while transparency independently enhances certain trust dimensions, it does not need to be coupled with human-like design features to be effective. This makes transparency an important factor in chatbot implementation, regardless of the context-dependent choices made for the visual and verbal design.

6. Implications

6.1 Theoretical implications

The findings from this study offer several contributions to the existing body of research on chatbot design, trust in AI, and human-computer interaction:

The results suggest that human-like visual design does not significantly impact trust or user experience, challenging prior assumptions in anthropomorphism theory. This suggests that the effectiveness of human-like design elements in chatbots may be more context-dependent than previously thought, aligning with the task-relevance perspectives. According to these, human-like elements are more effective when the interaction involves social or emotional engagement, as these cues help establish rapport and emotional connection (Gong, 2008). In contrast, for service chatbots focused on problem-solving or transactional tasks, users may prioritize efficiency and clarity over visual realism.

The study also highlights the distinction between emotional and cognitive trust components, as proposed by Mayer et al. (1995). Human-like verbal design was more effective at enhancing benevolence (an emotional dimension) than ability or integrity (cognitive dimensions), which suggests that verbal cues can evoke emotional trust without necessarily signaling competence or honesty. This expands trust theories by emphasizing the varying impact of different chatbot design elements on emotional vs. cognitive trust.

The significant effect of transparency on trust, especially in terms of ability and integrity, suggests that openly informing users they are interacting with a chatbot helps in building trust. However, transparency's lack of influence on perceived benevolence calls for further exploration into how users conceptualize honesty and fairness in non-human agents. This underscores the need to

further refine trust models in technology to account for the unique challenges of building integrity in non-human systems.

The absence of interaction effects between visual and verbal design can be explained by the dual-coding theory (Paivio, 1986), suggesting that users process these cues independently. This insight encourages researchers to explore how different cognitive channels function in tandem in human-computer interaction, with implications for both theory and the design of future studies.

6.2 Practical implications

The results have practical implications for the design of chatbots and AI systems, offering guidance for developers and organizations looking to improve user experience and trust:

The findings suggest that organizations should focus more on improving the chatbot's conversational abilities and functionality rather than investing in human-like visual design. Depending on the complexity of the task that the chatbots are supposed to perform, the level of humanness in the conversational style should be adjusted. Chatbots used for complex tasks should be designed with clear, concise communication, and less human-like language, to avoid confusion and enhance trust and user experience. Especially in contexts that require building rapport with users, service providers should prioritize using warm and empathetic language in chatbot interactions, since human-like verbal design improves perceived benevolence and emotional engagement. This is more important for applications, where emotional trust is paramount. Developers should recognize that anthropomorphic elements, particularly in visual and verbal design, are not universally effective. In functional, task-oriented contexts, efficiency and clarity may be more valuable than human-likeness. However, in social or emotional contexts, leveraging human-like verbal design can foster engagement and improve user experience. Understanding the task and context in which a chatbot will operate is essential for effective design.

Given the significant impact of transparency on trust, chatbot designers should ensure that the artificial nature of the chatbot is clearly communicated to users. Customers should be informed on the company's website that the chat assistant is a chatbot, not a human, prior to initiating the interaction. This could be done via a pop-up message or something similar, so that the user does not accidentally miss this information. This can help set appropriate expectations and reduce cognitive dissonance. Transparency can also serve as an ethical design practice, fostering a sense of openness and trustworthiness without needing to rely on human-like visual or verbal design.

In summary, the study's findings have important implications for both the theoretical understanding of chatbot design and practical strategies for improving trust and user experience. By focusing on transparency, verbal design, and task-relevant functionality, designers can create more effective and trustworthy chatbot interactions.

7. Limitations and future research

Although this study provides important insights, several limitations must be recognized. Firstly, participants in this study were not able to interact directly with the chatbot but instead watched a prerecorded interaction. This limitation may have influenced the findings, as passive observation might not fully capture the nuances of a real-time, dynamic interaction. Future research could build on this by allowing participants to engage directly with chatbots, which may lead to more accurate and comprehensive data on how human-like design and transparency affect user trust and experience in real-world settings.

Secondly, although efforts were made to gather a strong dataset, the final sample size ended up smaller than expected due to a significant number of missing data points. While the analysis was still based on 192 participants, which is sufficient for the statistical techniques applied, the reduced sample size might limit the generalizability of the results. A larger sample could have provided greater power to detect subtle effects or interactions that may have gone unnoticed in this study. Future research should aim for larger datasets to ensure the findings are both robust and widely applicable.

Thirdly, while the sample included participants from various nationalities, it was primarily composed of young and educated individuals, which could affect the generalizability of the outcomes.

On a more general level, another limiting factor of this study design is the consideration of gender. In this study, the chatbot was given a female identity, reflecting traits often stereotypically associated with women, such as warmth and empathy. However, it is important to critically examine these stereotypes. There is a notable gender bias in chatbot design, with around 78% of chatbots using female names, avatars, or descriptions (Feine et al., 2020). This bias is especially common in areas like customer service, branded communication, and sales (Feine et al., 2020). Given that bias should be minimized wherever possible, it is crucial to carefully evaluate and reflect on gender choices when designing and implementing chatbots. In this case, a female identity was chosen to align with the

expectations of users, where warmth and empathy are traits often connected to humanness. In future research, this gender bias could be looked at more closely by incorporating male and female chatbots, or potentially explore the effects of genderless chatbots.

Additionally, the study focused on specific aspects of trust and user experience, leaving out other potentially relevant factors, such as users' personality traits or previous experiences with chatbots. Future studies should consider these elements and investigate the long-term impacts of human-like design and transparency on user perceptions. Furthermore, qualitative methods could be employed to gain deeper insights into why users react to different chatbot designs in certain ways.

8. Conclusion

This thesis set out to explore how human-like design elements and transparency influence trust and user experience in service chatbots, aiming to bridge the gap between theory and practice in human-chatbot interaction design. The findings challenge assumptions about the uniform effectiveness of human-like design, revealing the nuanced roles of visual design, verbal design, and transparency. This highlights the importance of context in determining the impact of chatbot features. The insights on the design elements suggest that developers should prioritize conversational abilities over visual human-likeness, especially in task-oriented chatbots. Transparency practices can help manage user expectations and build trust. As AI continues to shape user interactions, understanding how design choices affect trust and user experience will be critical in creating ethical and user-centered technologies. This research serves as a step toward that goal, encouraging further exploration of how transparency and human-like features can be optimized in diverse applications.

References

Aldrup, K., Carstensen, B., & Klusmann, U. (2022). Is empathy the key to effective teaching? A systematic review of its association with teacher-student interactions and student outcomes. *Educational Psychology Review*, *34*(1), 1177–1216. https://doi.org/10.1007/s10648-021-09649-y

Ambalov, I. A. (2018). A meta-analysis of IT continuance: An evaluation of the expectation-confirmation model. *Telematics and Informatics*, *35*(6), 1561–1571. https://doi.org/10.1016/j.tele.2018.03.016

Anaya, L., Braizat, A., & Al-Ani, R. (2024). Implementing AI-based chatbot: Benefits and challenges. *Procedia Computer Science*, *239*, 1173–1179. https://doi.org/10.1016/j.procs.2024.06.284

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*(85), 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, *54*, 101473. https://doi.org/10.1016/j.tele.2020.101473

Bastiansen, M. H. A., Kroon, A. C., & Araujo, T. (2022). Female chatbots are helpful, male chatbots are competent? *Publizistik*, *67*. https://doi.org/10.1007/s11616-022-00762-8

Bejger, S., & Elster, S. (2020). Artificial intelligence in economic decision making: How to assure a trust? *Ekonomia I Prawo*, *19*(3), 411. https://doi.org/10.12775/eip.2020.028

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, *25*(3), 351–370. https://doi.org/10.2307/3250921

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/s15327965pli1302_01

Borau, S., Otterbring, T., Laporte, S., & Fosso Wamba, S. (2021). The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing*, *38*(7), 1052–1068. https://doi.org/10.1002/mar.21480

Bruckner, B. K. (2016). Eine Einführung in die Vertrauenstheorie. In *Organisationales Vertrauen initiieren. Entscheidungs- und Organisationstheorie.* (pp. 5–26). Springer Gabler. https://doi.org/10.1007/978-3-658-13492-1_2

Caldarini, G., Jaf, S., & McGarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, *13*(1), 41. mdpi. https://doi.org/10.3390/info13010041

Chattaraman, V., Kwon, W.-S., E. Gilbert, J., & Li, Y. (2014). Virtual shopping agents: Persona effects for older users. *Journal of Research in Interactive Marketing*, *8*(2), 144–162. https://doi.org/10.1108/jrim-08-2013-0054

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), 1–30. https://doi.org/10.1080/10447318.2020.1841438

Chen, F. F., Jing, Y., & Lee, J. M. (2014). The looks of a leader: Competent and trustworthy, but not dominant. *Journal of Experimental Social Psychology*, *51*, 27–33. https://doi.org/10.1016/j.jesp.2013.10.008

Chen, J. Y. C. (2022). Transparent human–agent communications. *International Journal of Human–Computer Interaction*, 1–2. https://doi.org/10.1080/10447318.2022.2120173

Chen, J., Guo, F., Ren, Z., Li, M., & Ham, J. (2023). Effects of anthropomorphic design cues of chatbots on users' perception and visual behaviors. *International Journal of Human–Computer Interaction*, *40*(14), 1–19. https://doi.org/10.1080/10447318.2023.2193514

Chen, J., Li, M., & Ham, J. (2024). Different dimensions of anthropomorphic design cues: How visual appearance and conversational style influence users' information disclosure tendency towards

chatbots. *International Journal of Human-Computer Studies*, *190*, 103320–103320. https://doi.org/10.1016/j.ijhcs.2024.103320

Choudhury, A., & Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis. *Journal of Medical Internet Research*, *25*(1), e47184. https://doi.org/10.2196/47184

Choung, H., David, P., & Ross, A. (2022). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, *39*(9), 1–13. https://doi.org/10.1080/10447318.2022.2050543

Crockett, K., Garratt, M., Latham, A., Colyer, E., & Goltz, S. (2020). Risk and trust perceptions of the public of artifical intelligence applications. *2020 International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/ijcnn48605.2020.9207654

Crockett, K., Goltz, S., Garratt, M., & Latham, A. (2019, June 1). Trust in computational intelligence systems: A case study in public perceptions. *2019 IEEE Congress on Evolutionary Computation (CEC)*. https://doi.org/10.1109/cec.2019.8790147

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

Dick, A., Chakravarti, D., & Biehal, G. (1990). Memory-Based inferences during consumer choice. *Journal of Consumer Research*, *17*(1), 82. https://doi.org/10.1086/208539

Doz, Y. (2011). Qualitative research for international business. *Journal of International Business Studies*, *42*(5), 582–590. https://doi.org/10.1057/jibs.2011.18

E te Pas, M., Rutten, W. G. M. M., Bouwman, R. A., & Buise, M. P. (2020). User experience of a chatbot questionnaire versus a regular computer questionnaire: Prospective comparative study. *JMIR Medical Informatics*, *8*(12). https://doi.org/10.2196/21982

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*(1), 275–298. https://doi.org/10.1146/annurev-psych-122216-011719

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender bias in chatbot design. *Chatbot Research and Design*, *11970*, 79–93. https://doi.org/10.1007/978-3-030-39540-7_6

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, *24*(4), 38–42. https://doi.org/10.1145/3085558

Garrett, J. J. (2011). *The elements of user experience: User-centered design for the web and beyond* (2nd ed.). New Riders.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, *24*(4), 1494–1509. https://doi.org/10.1016/j.chb.2007.05.007

Hassanein, K., & Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International Journal of Human-Computer Studies*, *65*(8), 689–708. https://doi.org/10.1016/j.ijhcs.2006.11.018

Hassenzahl, M. (2003). The thing and i: Understanding the relationship between user and product. *Funology*, *3*, 31–42. https://doi.org/10.1007/1-4020-2967-5_4

He, Y., Liu, C., & Luo, R. (2023). Emotional warmth and rejection parenting styles of grandparents/great grandparents and the social–emotional development of grandchildren/great grandchildren. *International Journal of Environmental Research and Public Health*, *20*(2), 1568. https://doi.org/10.3390/ijerph20021568

Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*(3), 245–250. https://doi.org/10.1016/j.chb.2015.02.026

Hoff, K. A., & Bashir, M. (2014). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

International Organization for Standardization. (2010). *Terms and definitions*. ISO 9241-210:2010(En) Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems. https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en

Jackson, R. B., & Williams, T. (2021). A theory of social agency for human-robot interaction. *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.687726

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of chatbots. *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*. https://doi.org/10.1145/3196709.3196735

Jeong, M., Minson, J., Yeomans, M., & Gino, F. (2019). Communicating with warmth in distributive negotiations is surprisingly counterproductive. *Management Science*, *65*(12). https://doi.org/10.1287/mnsc.2018.3199

Khurana, A., Alamzadeh, P., & Chilana, P. K. (2021, October 1). *ChatrEx: Designing explainable chatbot interfaces for enhancing usefulness, transparency, and trust*. IEEE Xplore. https://doi.org/10.1109/VL/HCC51201.2021.9576440

Kühne, R., & Peter, J. (2022). Anthropomorphism in human–robot interactions: a multidimensional conceptualization. *Communication Theory*, *33*(1). https://doi.org/10.1093/ct/qtac020

Kurpicz-Briki, M. (2023). *More than a chatbot: language models demystified.* Springer. https://doi.org/10.1007/978-3-031-37690-0

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *Lecture Notes in Computer Science*, *5298*, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J. (2009). Understanding, scoping and defining user experience. *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*. https://doi.org/10.1145/1518701.1518813

Lee, E., & Oh, S. Y. (2015). Effects of visual cues on social perceptions and self-categorization in computer-mediated communication. In *The Handbook of the Psychology of Communication Technology* (pp. 115–136). https://doi.org/10.1002/9781118426456.ch5

Lee, J. D., & See, K. A. (2004a). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Lee, J. D., & See, K. A. (2004b). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, *103*, 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69–96. Sagepub. https://doi.org/10.1509/jm.15.0420

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet*, *358*(9280), 483–488. https://doi.org/10.1016/s0140-6736(01)05627-6

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995a). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995b). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology. *ACM Transactions on Management Information Systems*, *2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

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

Melián-González, S., Gutiérrez-Taño, D., & Bulchand-Gidumal, J. (2019). Predicting the intentions to use chatbots for travel and tourism. *Current Issues in Tourism*, *24*(2), 1–19. https://doi.org/10.1080/13683500.2019.1706457

Meyer, J., Miller, C., Hancock, P., de Visser, E. J., & Dorneich, M. (2016). Politeness in machine-human and human-human interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 279–283. https://doi.org/10.1177/1541931213601064

Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2021). An emerging theory of avatar marketing. *Journal of Marketing*, *86*(1), 002224292199458. https://doi.org/10.1177/0022242921996646

Moorman, C., Zaltman, G., & Deshpande, R. (1992). Relationships between providers and users of market research: The dynamics of trust within and between organizations. *Journal of Marketing Research*, *29*(3), 314. https://doi.org/10.2307/3172742

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital Health*, *5*(5), 205520761987180. https://doi.org/10.1177/2055207619871808

Nass, C., & Moon, Y. (2000a). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., & Moon, Y. (2000b). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*. https://doi.org/10.1145/191666.191703

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *12*(5), 481–494. https://doi.org/10.1162/105474603322761289

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford Univ. Pr.

Parboteeah, D. V., Valacich, J. S., & Wells, J. D. (2009). The influence of website characteristics on a consumer's urge to buy impulsively. *Information Systems Research*, *20*(1), 60–78. https://doi.org/10.1287/isre.1070.0157

Pawlik, V. P. (2022). Design matters! How visual gendered anthropomorphic design cues moderate the determinants of the behavioral intention towards using chatbots. *Chatbot Research and Design*, *13171*, 192–208. https://doi.org/10.1007/978-3-030-94890-0_12

Pitardi, V., & Marriott, H. R. (2021). Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology & Marketing*, *38*(4), 626–642. https://doi.org/10.1002/mar.21457

Plantak Vukovac, D., Horvat, A., & Čižmešija, A. (2021). Usability and user experience of a chat application with integrated educational chatbot functionalities. *Learning and Collaboration Technologies: Games and Virtual Environments for Learning*, 216–229. https://doi.org/10.1007/978-3-030-77943-6_14

Powderly, H. (2024, March 21). *Guide to AI chatbots for marketing: Options, capabilities, and tactics to explore*. EMARKETER. https://www.emarketer.com/insights/chatbot-market-stats-trends/

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, *25*(4), 145–182. https://doi.org/10.2753/mis0742-1222250405

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, *151*, 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Camebridge University Press.

Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior*, *126*, 107034. https://doi.org/10.1016/j.chb.2021.107034

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

Roy, R., & Naidoo, V. (2021). Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, *126*, 23–34. https://doi.org/10.1016/j.jbusres.2020.12.051

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (1996). Organizational trust: Philosophical perspectives and conceptual definitions. *The Academy of Management Review*, *21*(2), 337–340. https://www.jstor.org/stable/258662

Schrepp, M., Hinderks, A., & ThomaschewskiJ. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(6), 103. https://doi.org/10.9781/ijimai.2017.09.001

Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2017). When do we need a human? Anthropomorphic design and trustworthiness of conversational agents. In *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS* (Vol 10). AISeL.

Smestad, T. L., & Volden, F. (2019). Chatbot personalities matters. *Internet Science*, *11551*, 170–181. https://doi.org/10.1007/978-3-030-17705-8_15

Song, S. W., & Shin, M. (2022). Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: The moderating role of avatar familiarity. *International Journal of Human–Computer Interaction*, 1–16. https://doi.org/10.1080/10447318.2022.2121038

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, *19*(3), 529–545. https://doi.org/10.1111/jcc4.12066

Wang, L. C., Baker, J., Wagner, J. A., & Wakefield, K. (2007). Can a retail web site be social? *Journal of Marketing*, *71*(3), 143–157. https://doi.org/10.1509/jmkg.71.3.143

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. https://doi.org/10.1145/3025453.3025496

Xu, Y., Bradford, N., & Garg, R. (2023). Transparency enhances positive perceptions of social artificial intelligence. *Human Behavior and Emerging Technologies*, *2023*. https://doi.org/10.1155/2023/5550418

Appendix A

Pre-test material

Interview introduction and consent

Thank you for agreeing to participate in this research study. The goal of this pre-test is to identify which visual and verbal manipulations most effectively achieve the desired perceptions of human-likeness and roboticness in the design of a chatbot. Your feedback is crucial in helping me refine these designs before conducting my primary study.

During this interview, I will show you different visual representations of a chatbot and show you scripts of chatbot interactions featuring two distinct styles of chatbot dialogue. I will ask you to share your thoughts and perceptions on these elements. Your insights will help me understand which designs and dialogues are most effective in conveying human-like or robotic traits.

I would like to notes during our conversation to accurately capture your responses.

Do I have your consent to proceed?

Table A1

Pre-test interview questions

Topic	Questions
Visual design	1. Looking at these avatars, which one do you perceive as the most human-like? Why? 2. Looking at these avatars, which one do you perceive as the most robotic? Why? 3. Considering the chatbot context, which avatar do you think fits best as a chatbot? What makes you feel this way? 4. What suggestions do you have for improving the chatbot's visual design to better convey the desired human-like or robotic traits?
Verbal design	1. Can you describe the differences between the two scripts you read? 2. How would you describe the chatbot's verbal style in the human-like design? 3. How would you describe the chatbot's verbal style in the robotic design? 4. What suggestions do you have for improving the chatbot's verbal interactions to better convey the desired human-like or robotic traits?
General	1. Is there anything else you want to share or suggest for improvement of my research? Anything that stood out to you?

Table A2

Human-like chatbot script

Speaker	Text
User	Hi, I'm having trouble with my service. Can you help me?
Chatbot	Hello! I'm Anna and I'm here to help you. Could you please provide more details about the issue you're facing?
User	Sure. My internet connection has been really slow for the past few days.
Chatbot	I'm sorry to hear that. Let's see what we can do to improve your connection. Can you confirm if you're experiencing this issue on all your devices or just one?
User	It's happening on all my devices.
Chatbot	Thank you for the information. I'll run a quick check on your service status. This might take a moment. [Pause]
Chatbot	I've checked your service status, and it looks like there might be an issue in your area. I'll need to reset your connection from my end. This will take a few minutes. Is it okay if I proceed?
User	Yes, please go ahead.
Chatbot	Great! I'm starting the reset now. Please hold on for a moment. [Pause]
Chatbot	The reset is complete. Could you please restart your router and check if the connection has improved?
User	It seems to be working better now.
Chatbot	That's wonderful to hear! If you encounter any more issues, feel free to reach out. Is there anything else I can assist you with today?
User	No, that's all. Thank you!
Chatbot	You're welcome! Have a great day!

Table A3

Not human-like chatbot script

Speaker	Text
User	Hi, I'm having trouble with my service. Can you help me?
Chatbot	Provide details of your issue.
User	Sure. My internet connection has been really slow for the past few days.
Chatbot	Acknowledged. Is this issue on all devices or just one?
User	It's happening on all my devices.
Chatbot	Understood. Checking service status. This will take a moment.
	[Pause]
Chatbot	Service status indicates a possible issue in your area. I will reset your connection. This will take a few minutes. Can I proceed?
User	Yes, please go ahead.
Chatbot	Resetting. Wait.
	[Pause]
Chatbot	Reset complete. Restart your router and check your connection.
User	It seems to be working better now.
Chatbot	Confirmed. Connection improved. If further issues arise, contact us again. Do you need further assistance?
User	No, that's all. Thank you!
Chatbot	Acknowledged. Goodbye.

Appendix B

Primary study material

Table B1

Demographic questions

Variable	Options/Format
Age	What is your age? (Open-ended, e.g., 25)
Educational level	What is your highest achieved level of education? ● High school ● Bachelor's degree ● Master's degree ● Doctorate ● Other
Nationality	What is your nationality? ● German ● Dutch ● Other

Table B2

Trustworthiness scale

Variable	Definition	Items
Ability	Ability refers to the chatbot's ability to perform its tasks effectively.	1. The chat assistant is very capable of performing its tasks. 2. The chat assistant successfully accomplishes what it sets out to do. 3. The chat assistant has much knowledge about the tasks it needs to perform. 4. I feel very confident about the chat assistant's capabilities. 5. The chat assistant has specialized features that can improve my experience. 6. The chat assistant is well-designed to handle the tasks it was used for.
Benevolence	Benevolence pertains to the chatbot's willingness and ability to assist users.	1. The chat assistant is very concerned about my needs. 2. The chat assistant prioritizes my questions and concerns. 3. The chat assistant would not knowingly provide unhelpful information. 4. The chat assistant looks out for what is important to me. 5. The chat assistant goes out of its way to help me.
Integrity	Integrity refers to the consistency and dependability of the chatbot's actions.	1. The chat assistant has a strong sense of reliability. 2. I never have to wonder whether the chat assistant will perform as expected. 3. The chat assistant consistently provides accurate information. 4. The chat assistant's actions and responses are very consistent. 5. I trust the chat assistant's underlying principles. 6. The chat assistant operates in a manner that is consistent and dependable.

Note. These items were adapted from a scale measuring ability, benevolence, and integrity by Schoorman et al. (1996).

Table B3

User experience questionnaire

Variable	Definition	Items
Attractiveness	Measures the overall appeal and pleasantness of interacting with the chatbot.	Annoying - enjoyable Bad - good Unlikable - pleasing Unpleasant - pleasant Unattractive - attractive Unfriendly - friendly
Perspicuity	Assesses how easy the chatbot is to understand and interact with.	Not understandable - understandable Difficult to learn - easy to learn Complicated - easy Confusing - clear
Novelty	Evaluates the creativity and innovation of the chatbot's features.	Dull - creative Conventional - inventive Usual - leading edge Conservative - innovative
Stimulation	Measures the chatbot's ability to engage and motivate the user.	Inferior - valuable Boring - exciting Not interesting - interesting Demotivating - Motivating
Dependability	Assesses the reliability and trustworthiness of the chatbot.	Unpredictable - predictable Obstructive - supportive Not secure - secure Does not meet expectations - meets expectations
Efficiency	Evaluates how quickly and effectively the chatbot performs its tasks.	Slow - fast Inefficient - efficient Impractical - practical Cluttered - organized

Note. These items were taken from the User Experience Questionnaire developed by Laugwitz et al. (2008). An example of this scale is illustrated in Figure B1.

Figure B1

Example of UEQ item

Annoying ① ② ③ ④ ⑤ ⑥ ⑦ Enjoyable

Appendix C

Statement of AI use

In this study, the author utilized ChatGPT-4o, developed by OpenAI, to enhance the quality of the language and to assist in the statistical analysis and coding processes. While ChatGPT-4o provided valuable support, the author takes full responsibility for all content, interpretations, and any errors within this work.