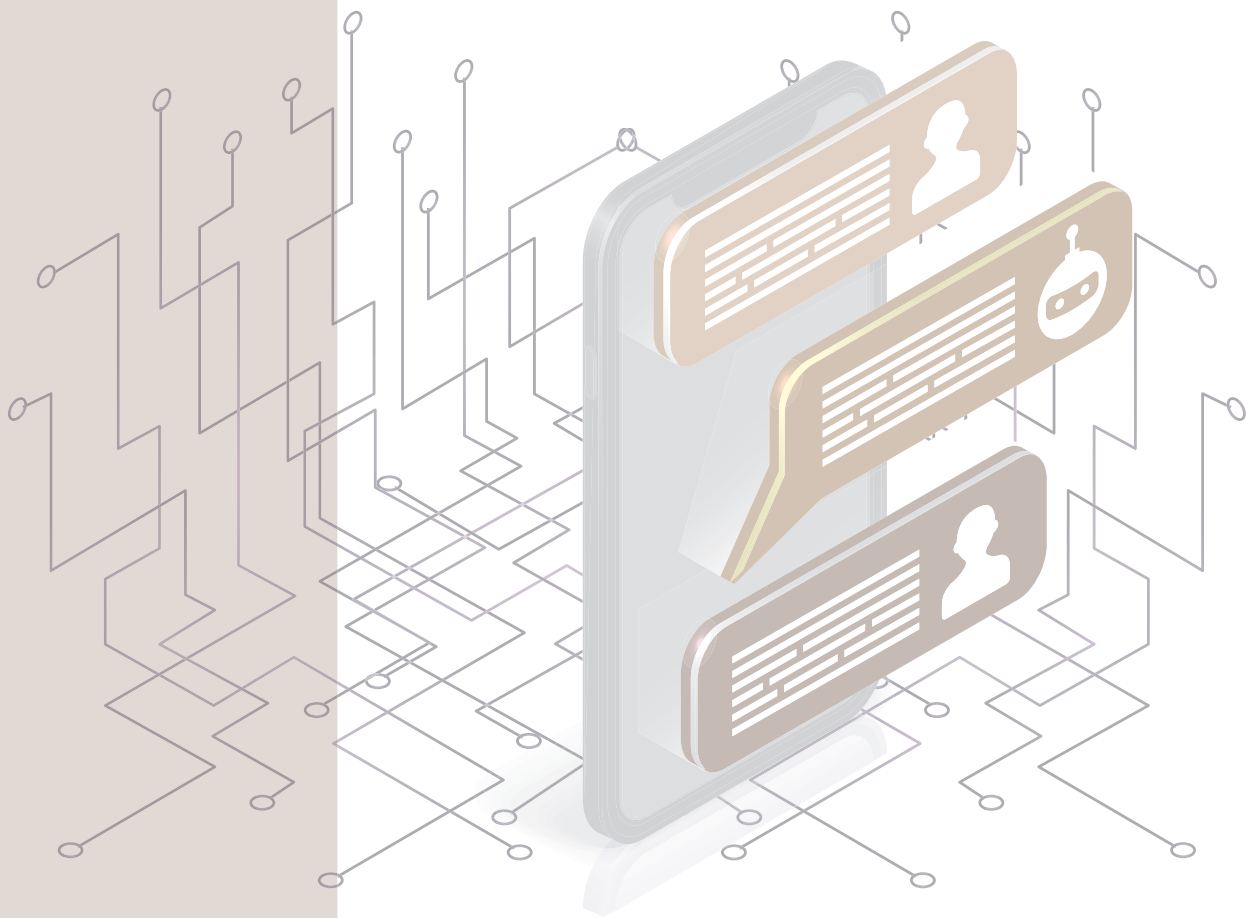# Designing for a better user experience

The effects of visual appearance, chatbot gender and context on perceptions of trust, perceived ease of use, empathy, customer satisfaction, competence, social presence, and intention to use

Leandra Valeria Simbach
s1828592
Master Thesis
University of Twente

Supervisors:
Dr. Joyce Karreman
Dr. Thomas van Rompay

Abstract

Research objective: Chatbots are becoming increasingly popular as service assistants in a time where users have an increased urge for quick responses. Chatbots can be implemented in various contexts and bring along many advantages for users as well as for the companies integrating them. The majority of chatbots applied are portrayed with the help of an avatar but not all are equally effective. This study investigates different chatbot designs aiming to demonstrate whether visual appearance of the chatbot, gender of the chatbot, or the context the chatbot is administered in have an influence on users' perceptions of the chatbot regarding customer satisfaction, perceived ease of use, trust, empathy, social presence, competence, and intention to use.

Method: A quantitative study with a 2x2x2 design was conducted. Visual appearance has been compared with the help of a smiling and a non-smiling facial expression, gender has been examined using a male and a female version, and a healthcare versus a financial context have been opposed. A survey with 232 participants was conducted. Previously, a pre-test to validate the designs was undertaken.

Findings: The context the chatbot interacts in did produce the most significant effects. For most dependent variables, there was a difference in the scores of contexts. In general, the healthcare context produced higher numbers, with a few exceptions. The visual appearance as well as the gender of the chatbot are only relevant the interaction with each other or in relation to context. There are no effects of any variable on perceived ease of use.

Conclusion: The appearance of a chatbot, their gender, as well as the context of interaction all influence perceptions of the chatbot, with context having the main influence on perceptions of chatbot design. The study provides implications for the integration of social presence theory and user-centred design.

Key words: chatbot design, gender, context, visual appearance, facial expression, human-robot interaction

Table of contents

1. Introduction

In the digital age, the rise of artificial intelligence (AI) and machine learning has revolutionized how businesses and individuals interact, with chatbots emerging as one of the most significant applications of these technologies. People's need for fast and goal-oriented responses is increasing, meaning that they want to achieve a specific objective and receive a practical solution. Users request customized answers, and every company wants to offer the best experience possible. Caused by a lack of skilled workers and the loans these people have to be paid, as well as consumers increased wish for quick responses, the search for an alternative for human employees is comprehensible. Here, the use of a chatbot can support the companies as well customers.

Chatbots are software programs designed to simulate human-like conversations, responding to user inputs through text or voice in real-time. IBM (n.d.) depicts "A chatbot is a computer program that simulates human conversation with an end user". Chatbots have rapidly evolved from simple rule-based systems to sophisticated AI-driven tools capable of natural language understanding and context awareness. This evolution has made chatbots an indispensable component in various industries.

The first chatbots have been introduced in the 1960's (Radziwill & Benton, 2017) and the area of chatbot design is increasing rapidly and more companies are applying the use of a chatbot as a service agent on their website. Here, ensuring that the chatbot is perceived as helpful to the customer and that the interaction is perceived as pleasant is important (Chandra et al., 2022). Given that a chatbot is being accepted by the user, they can serve as digital companions that bridge communication gaps, offer support and provide assistance.

In recent years, chatbots have emerged as a transformative tool in the digital landscape, revolutionizing the way businesses, organizations, and individuals interact with technology. As technology continues to advance, the role of chatbots is expected to expand further, making them a vital component in the digital strategies of businesses and institutions across the globe but not every conversation is comparable. There are several reasons why a user may contact a chatbot. The chatbot's ability to answer users' questions to a full extent was the most important aspect for users to use a chatbot in a study by Islind et al. (2023).

Chatbots are suitable to be applied in many domains such as education, customer service or simple amusement (Shawar & Atwell, 2007). Moreover, the advantage of chatbot responsibilities is that they are available all around the clock (Thomas & Fowler, 2021), which is especially useful with regard to fast responses and service on the weekend. During peak times (e.g., product launches or holiday seasons), chatbots can manage increased demand without sacrificing quality, leading to operational cost savings and improved service (Janssen et al., 2022). Users are less likely to become frustrated with long hold times or delayed responses, which improves the overall experience. Additionally, a chatbot can access tons of data at the same time and is provided with all information available on the internet.

One of the primary reasons for the growing importance of chatbots is their ability to provide instant and personalized communication. In an era where customers expect immediate responses and seamless experiences, chatbots can handle a vast volume of inquiries simultaneously, ensuring that

users receive timely support without waiting in queues. By instantly answering queries and requests without requiring human assistance or manual research, chatbots can simplify the process of finding information for users (IBM, n.d.). This capability not only enhances customer satisfaction but also improves operational efficiency by reducing the workload on human agents. For businesses, this translates into significant cost savings (Wang et al., 2023). In 2024, 58% of companies worldwide have already integrated some version of a chatbot in their website (Cherniak, 2024) and 60.02% of the people aged between 25 and 34 reach out to a chatbot on a regular basis, mainly for educational purposes (Statista, 2024)

Furthermore, the importance of chatbots extends beyond customer-facing applications. In healthcare, for example, chatbots are being utilized to provide mental health support, offer preliminary diagnoses, and manage patient appointments, making healthcare services more accessible and efficient. In education, chatbots can serve as virtual tutors, helping students with their queries and offering personalized learning experiences. These diverse applications demonstrate the versatility and potential of chatbots to transform traditional service models and create new opportunities for innovation.

Chatbots have become a crucial asset in the modern world due to their ability to enhance communication, improve accessibility, reduce costs, and provide valuable insights through data analytics. By offering users a non-judgmental, always-available source of assistance, chatbots can improve access to mental health resources and reduce the stigma associated with seeking help (Kiuchi et al., 2023). Overall, engaging with chatbots can make information gathering more efficient, convenient, and satisfying. By keeping users engaged, chatbots can improve the chances of conversion (e.g., completing a sale or registration) and ensure users feel more connected to the service or brand.

Nevertheless, there are some obstacles for the implementation of successful chatbots. Belanche et al. (2023) express that users do not stay loyal to a robotic advisor, as they lack a sense of feeling connected to them as a personal assistant. Moreover, many simple chatbots have non-user-friendly interfaces and users find it disconcerting to engage in an unpleasant environment (Kumar & Ali, 2020). Abdellatif et al. (2020) support this notion and add that people are hesitant to use a chatbot when it feels unnatural.

The practical deployment of chatbots across various sectors has illuminated both their immense potential and the challenges they face in real-world applications. As digital assistants, chatbots have significantly transformed how organizations engage with customers, handle inquiries, and manage operations. However, the success of chatbots in practice depends on several factors, including their design, integration, and user acceptance (Janssen et al., 2022). The design implications differ per domain (Linder, 2020). Different situations may require different adaptions of the chatbots visual and characteristic appearance. Linder (2020) explains that a travel chatbot must be designed with cultural sensitivity, a healthcare chatbot should be equipped with medical terminology, and in finance, a chatbot needs to guide users through complex transactions.

This study will investigate how different characteristics incorporated in a chatbot shape perceptions of a chatbots' gender, its visual appearance and the context it operates in, moreover,

how perceptions change when these variables interact with each other. Visual appearance has been selected as it is the first thing that users notice and it significantly shapes a chatbots personality, especially if a static avatar is used without animation. The look of an avatar can decide whether users believe it to be helpful, similar to judging if a person is skilled and of help (Smestad, 2018). According to Chung and Lee (2020), the appearance is even more important when the chatbot is intended to portray an identity. Gender is chosen because it relates to social norms and biases and can influence how users rate different abilities. Based on culture or nurture, people associate different tasks to one or the other gender (Aumüller et al., 2024). Lastly, the context in which a chatbot operates determines how the chatbot should behave, whether it should mainly focus on functionality or if confidentiality and trustworthiness are important as well (Følstad et al., 2019). These factors are key characteristics that shape chatbot perceptions.

2. Theoretical framework

By aligning chatbot personalities with the customers' expectations, businesses can create more personalised and engaging experiences, leading to higher levels of customer engagement (Lannoy, 2017). There are many facets that should be considered when designing a chatbot, for example, trust, customer satisfaction, and empathy perceived by the user, influencing how strong the engagement with the chatbot will be. The perceptions of these variables can, however, vary depending on the characteristics of the chatbot, which in this study will be the gender of the chatbot, the visual appearance of the chatbot, and the context in which the interaction will take place.

2.1 Factors influencing chatbot perception

It is proven that, when talking to a chatbot, users behave according to the same social standards as when talking to another human (Baxter et al., 2018). As research has shown, there are differences in preferences, behaviours, and satisfaction levels which can also vary depending on the industry, product, or service. People are using a chatbot for convenience and with regard to the aspect of saving time (Hari et al., 2022). However, not one standardised chatbot can satisfy all users' needs, demands and wishes.

2.1.1 Visual appearance

The appearance of the chatbot is pivotal for a trustworthy and pleasant impression. The design features of the chatbot outline an essential factor in the overall experience of the service (Islind et al., 2023). Zierau et al. (2020) suggest that using human features can remarkably influence users' perceptions of social presence, which in turn enhances users' trust in the chatbot. These cues can be gestures as well as a confirmation of understanding. The behaviour of many chatbots is still machine-like. However, subtle facial expressions in robots can impact human judgments of them as social counterparts (Mathur & Reichling, 2009). Mathur and Reichling (2009) state that it is enough if these facial expressions are slightly raised corners of the mouth. In their study from 2009, they found that a slight display of emotion can positively affect perceptions of likability. For theses subtle cues to work it is enough to minimally articulate a smile.

Expectations about the interaction are also derived from the appearance of the chatbot. (Go & Sundar, 2019). A study by Krumhuber et al. (2007) suggests that facial dynamics play an important role in social interactions, particularly in situations involving trust and cooperation. In the study by Krumhuber et al. (2007), an authentic smile, a fake smile and a neutral facial expression have been compared. The results show that participants did notice the changes in the chatbots' facial expression. Therefore, these social cues should be implemented in the design of chatbots as well.

The appearance of the chatbot does not only influence liking, but it also hints at the expected satisfaction with the outcome of the interaction which in turn positively influences user experience. The level of humanness seems to predict the perceived competence of an avatar (Miao et al., 2022). The appearance of a chatbot leads to perceptions of the character (Laakasuo et al., 2021), hence, this may as well impact the perceived level of trust towards the digital personality. Also, the self-

presentation of the chatbot matters, as it communicates competence (Følstad et al., 2018), and in turn leads to a higher perception of usefulness.

Friendliness is especially important when users are asked to disclose about sensitive topics (Bhakta et al., 2014). A friendly chatbot can enhance user engagement by creating a more welcoming and approachable interaction environment. Users are more likely to interact with a chatbot that exhibits friendly behaviour and tone (Brandtzaeg & Følstad, 2017). Additionally, a friendly chatbot can enhance the overall user experience by creating a more pleasant and enjoyable interaction. The psychological distance between the chatbot and the user becomes smaller when the chatbot is friendly and welcoming (Zogaj et al., 2023) and users are more likely to return to a chatbot that they perceive as friendly and easy to communicate with.

In this study, the differences in perception of the visual appearance of the chatbot are investigated using a friendly visual appearance demonstrating a more humane character arousing warmth compared to a chatbot with a neutral facial expression showing no emotionality. It is expected that the smiling chatbot will lead to better results than the chatbot with the neutral facial expression.

2.1.2 Gender

Similar to encounters with a real human, gender can also have an influence on the interaction with a virtual agent (Bastiansen et al., 2022). Small cues on gender can influence the behaviour and, therefore, the reactions of the users. In 2023, only 37% of chatbots have been female (Statista, 2023). Nevertheless, the most famous ones', namely Alexa, Siri, and Cortana, are all female.

Gender stereotypes and role congruity theory suggest that individuals may have expectations about how gender aligns with specific roles or tasks (Amin et al., 2014). Amin et al. (2014) say that stereotypes still play a major role in how we perceive different genders. In their study, with regard to role congruity theory, they point out that women are associated as caregivers and men expected to be the provider. Due to cultural norms, biases, and experience, people may have a general preference of one gender over the other (Schillaci et al., 2024). Subsequently, users transfer their gender specific stereotypes to virtual personalities (Brahnam & De Angeli, 2012). Brahnam and De Angeli (2012) rationalize that users tend to replicate social interaction patterns they use with humans when engaging with virtual agents.

Some people express the wish to be treated by a woman in healthcare. This is caused by the attributed warmth and higher ability for compassion (Borau et al., 2021). Besides, Toader et al. (2019) reported that female virtual assistants are more commonly forgiven for errors compared to male chatbots. Moreover, in the study by Toader et al. (2019), it was discovered that participants are more willing to disclose sensitive information to a female chatbot, compared to a male chatbot. Also, Toader et al. (2019) encountered that female virtual assistants lead to more positive consumer responses. However, women, more often than men, are judged by their attractiveness (Brahnam & De Angeli, 2012). This may lead to a bias about the competencies of the virtual character as well.

The UNESCO has warned that discussions about failures of personal assistants such as 'Alexa'

may lead to gender stereotypes. Therefore, it seems relevant to explore whether a male chatbot leads to positive user experiences as well and may be able to replace or at least support female virtual agents (Feine et al., 2020). These findings give reason to assume that the gender of the chatbot may have an influence in specific contexts. Kim et al. wrote in 2021 that female chatbots are expected for counselling and in healthcare, while male chatbots are associated with professional work and customer service. This supports the idea that there are indeed conventional expectations.

To validate whether these findings apply to different contexts, in this study, a male and a female chatbot are tested. There is no gender neutral chatbot included as non-binary chatbots are very scarce and barely used. It is expected that the female chatbot will be perceived as more empathic and more advantageous to be applied in a healthcare context, while the male chatbot will be associated with more competence and better suitable for the financial context.

2.1.3 Context

In general, whether a person wants to use a chatbot is dependent on the context and the personal involvement (Brachten et al., 2021). According to Brachten et al. (2021), the complexity of a task and the perceived value of the interaction are motivational factors deciding if a chatbot can be useful. Chatbots are also valuable and, in some places even essential, as they provide faster support than a human can. Different contexts and differences in severity of a topic or situation, however, may require different adaptations of chatbots. While one requirement for a customer service chatbot is a recommendation on how to use a product, another situation including a social companion chatbot requires emotional support and understanding. The two contexts compared in this study are healthcare and finance. These two contexts have been selected as the primary purposes of these two domains differ very much. In finances, the primary function of a chatbot is to be a consultant. The main purpose in this case is to simplify complex processes and give precise and accurate responses. In healthcare, a chatbot may need to track symptoms, provide patient care and act as a therapeutic tool. Here, the main purpose is understanding health-related issues and provide compassionate support.

The COVID-19 pandemic has increased the need for psychological support and comforting care (Torous et al., 2021), as human touch was not available anymore. This has raised awareness for the gains of a medical chatbot. Implementing chatbots in healthcare settings can help patients get a better understanding of the problems they are facing, reduce the burden on hospitals when there are many patients to handle, and save time for doctors (Kandpal et al., 2020). Chatbots can benefit patients by providing them with quick and easy access to information, personalised support, and round-the clock assistance in the absence of physicians (Vanimireddy et al., 2023). Overall, chatbots have been used effectively in healthcare to provide more understanding of diseases, reduce exposure to infections, assist workers with rapid and accurate solutions, and communicate with doctors.

In finances, chatbots provide round-the-clock customer support, allowing users to access banking services and assistance at any time of the day, even outside of traditional banking hours.

Implementing chatbots can help financial institutions reduce operational costs by automating routine tasks and customer service interactions, leading to efficiency gains and cost savings, and chatbots streamline financial transactions such as fund transfers, bill payments, account inquiries, and investment management, making banking processes more efficient and convenient for users (Alt et al., 2021). Chatbots offer a range of benefits for both financial institutions and consumers.

Regarding the two different contexts which will be tested in this study, it is expected that empathy and commitment to the user will be more important in the healthcare context, while competence will play a bigger role in the financial context.

2.2 Key variables of interest

As mentioned above, a lot of factors can influence the perceptions a user has about a chatbot. As there is a multiplicity of characteristics attributed to chatbots, seven variables specifically considered to be important in this study will be elaborated. These are customer satisfaction, trust, empathy, perceived ease of use, social presence, competence, and intention to use.

2.2.1 Customer satisfaction

Customer satisfaction is the main indicator by which a customers' experience can be evaluated. The satisfaction with the chatbot as well as a positive experience with the chatbot are important factors that define whether a user will use a chatbot repeatedly and if the user rates the interaction as helpful. Moreover, factors such as a human-like character and friendliness contribute to the aspect of customer satisfaction. Satisfaction can be enhanced by providing a pleasant design (Amin et al., 2014). In this study by Amin et al. (2014), an aesthetic appeal and intuitive design are the most important implications for chatbot implementation.

By designing chatbots with engaging and personable personalities, businesses can enhance customer satisfaction levels (de Lannoy, 2017). The experience with a chatbot is significantly enhanced if the chatbot is fun to interact with (Islind et al., 2023). Expectations about the satisfaction with the interaction are also derived from the appearance of the chatbot (Go & Sundar, 2019). Adding to this, Ekinci and Dawes (2009) are cited in an article by de Haan (2018) stating that smiling is a customer-oriented behaviour and increases customer satisfaction.

Further, a study by Sanjeed et al. (2020) showed that a male banking chatbot increases customer satisfaction significantly more than a female banking chatbot. In their study, Sanjeed et al. (2020) applied a 2x2 design comparing chatbot gender and the congruence of the chatbot gender with the respondents' gender. In the interaction with a congruent gender as well as with an incongruent gender, the male chatbot showed better results than the female chatbot. This is explained by the fact that a male person may be expected in the context of financial advice and there may be preconceived notions that a man is the best fit for this task.

On the other hand, based on stereotypes and social norms, it is assumed that this same effect may be generated for a woman in the medical context, as women are more commonly associated with healthcare. Combining the aspect of gender with the aspect of appearance, it is expected that in both contexts, a smiling chatbot will lead to a better customer satisfaction than a neutral chatbot.

H1: A smiling chatbot, compared to a neutral chatbot, will lead to a better customer satisfaction.

H2: In a financial context, a male banking chatbot will lead to a higher customer satisfaction, while in a healthcare context, a female healthcare chatbot will lead to a higher customer satisfaction.

H3: In healthcare context, a smiling female chatbot, compared to a neutral male chatbot, will lead to a higher customer satisfaction.

2.2.2 Perceived ease of use

Perceived ease of use is a critical factor influencing users' experiences with digital services. Perceived ease of use refers to the degree to which users believe that using a system will be free of effort (Amin et al., 2014). A friendly chatbot, intuitive design, and efficient functionality make it easier for customers to perform transactions, and navigate the platform, leading to a positive perception of the service (Anugrah, 2020). The navigation through the chatbot should be intuitive, the conversation should flow naturally, and it should be made easy for users to communicate their preferences (Kasilingam, 2020).

Social factors, such as peer influence, can shape attitudes towards the chatbot and thereby diminish the perceived ease of use (Jin & Eastin, 2023). A topic associated with feminine traits, such as healthcare, may therefore impact the ease of use perceived with gender-untypical chatbots. Similarly, the perceived ease of use in the financial context may be higher with a male chatbot, as finances are associated to be a male domain. The smoothness of an interaction is evaluated more positively when the experience is congruent with expectations. This means, as a woman is commonly expected in a medical context, a reality differing from this convention is perceived less well, as it contradicts the users' habits. Conditions differing from expectations create dissonance.

H4: A smiling chatbot, compared to a neutral chatbot, will lead to higher perceived ease of use.

H5: In a healthcare context, a female chatbot will lead to higher perceived ease of use, while in a financial context, a male chatbot will lead to higher perceived ease of use.

2.2.3 Trust

Trust in the conversational partner is a precondition for a truthful cooperation in any interaction where one character needs to open up to another character. Trusting someone means accepting a risk (Paradeda et al., 2016). Mayer et al. (1995) defined 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". Likewise, trust is a prerequisite for people to engage with electronic services (Beldad et al., 2010). Trust in a chatbot is essential for the user to be willing to engage in the interaction. The concept of trust has been investigated for decades in many fields, particularly psychology and social science (Paradeda et al., 2016). Trust in customer service is related to factors such as the chatbots' ability to correctly interpret user requests and provide useful

answers, as well as its human-likeness, self-presentation, and professional appearance (Nordheim et al., 2019).

Regarding the physical appearance of chatbots, humanoid appearance can contribute to evoking trust in the user (Gebhard et al., 2021). The behaviour of many chatbots is still machine-like. This diminishes trust, as users look for cues of behaviour in their interaction partners, even in case the counterpart is a computer (Brave & Nass, 2007). Subtle facial expressions in robots can impact human judgments of their trustworthiness as social counterparts (Mathur & Reichling, 2009). A study by Krumhuber et al. (2007) suggests that facial expressions play an important role in social interactions, particularly in situations involving trust and cooperation. In the study by Krumhuber et al. (2007), levels of trust for a smiling and a non-smiling chatbot have been compared and the willingness to cooperate was significantly higher for the smiling chatbot. They demonstrated that counterparts who show an authentic smile are perceived as more likeable, attractive, and trustworthy than those who showed a neutral expression.

Similarly, Zierau et al. (2020) suggest that using human features can remarkably influence users' perceptions of engagement, which in turn enhances users' trust in the chatbot. Facial dynamics serve as a cue to infer trustworthiness of others. These cues can be gestures as well as a confirmation of understanding. The appearance of a chatbot also leads to perceptions of the character (Laakasuo et al., 2021), hence, this may as well impact the perceived level of trust towards the digital personality.

As trust is the belief or confidence that one can rely on another person's actions, words, or decisions (Opheim, 2023), in healthcare, trust is essential for building a strong patient-provider relationship based on mutual respect, honesty, and reliability. Patients need to trust their healthcare providers to make informed decisions about their health and well-being (Opheim, 2023).

Finances are another context where trust and a positive consumer experience are crucial. According to Beldad et al. (2010), people calculate the risks of financial investments more cautiously than situations that do not include monetary transactions. This means that trust must be very high to be willing to engage in a financial context. The number of chatbots implemented in finances is increasing but adoption and usage numbers are still low (Sugumar & Chandra, 2021). Furthermore, research about chatbots in financial contexts is scarce. A study by Ng et al. (2020) revealed that the presence of a chatbot did have a significant positive effect on the perceived level of trust in a financial decision. Financial advisors and institutions can build trust by showing genuine concern for their clients' needs, providing transparent and accurate information, and offering personalised services. In finances, humanness factors such as facial expression are also crucial for chatbots adoption and display various features (Suguman & Chandra, 2021). This accounts for healthcare as well. Again, these perceptions of humanness can be increased by an authentic smile. There are no studies providing relevant insight into the preferred gender of a chatbot in the financial sector, although most chatbots are female.

H6: A smiling facial expression, compared to a neutral facial expression, will lead to more trust.

2.2.4 Empathy

Empathy is a decisive component for an uninhibited conversation between a chatbot and the user. Empathy is essential for understanding clients' and patients' perspectives, needs, and concerns and tailoring services accordingly (Baldner et al., 2015). Agents that can understand the context of a situation and respond appropriately can evoke empathy in humans (Paiva et al., 2017). By considering social and cultural aspects of interactions, agents can create empathy in humans. Empathy can be portrayed in responding to personal needs and not only providing standardised answers (Gebhard et al., 2021).

Even though users are aware that they are talking to a computer, they treat virtual agents as social entities. Likewise, emotions expressed by virtual agents influence users' perceptions in a positive manner (Liu & Sundar, 2018). Artificial agents can display emotions through facial expressions, gestures and tone of voice, which can be seen as empathic behaviour from humans who perceive these as emotional signals (Paiva et al., 2017). Agents can provide empathic feedback by recognising and responding to the emotional states of users. For example, an agent may show empathy towards a user experiencing distress or offer encouragement to a user in a neutral emotional state (Paiva et al., 2017). Further, the chatbot could show compassion for a personal situation.

Empathy and the evoked feeling of caring are crucial for sharing health-related as well as finance-related information in real-life and therefore need to be transferred to chatbot design as well (Bhakta et al., 2014). Empathy promotes ethical behaviour and cooperation in financial transactions (Baldner et al., 2015). In finances, empathy is essential for understanding clients' perspectives, needs, and concerns and tailoring services accordingly (Baldner et al., 2015). In healthcare settings, empathy involves healthcare providers being able to understand and connect with their patients on an emotional level, showing compassion, and demonstrating that they care about the patient's well-being (Opheim, 2023). Bickmore et al. (2009) confirm that, in a healthcare context, empathy is indispensable to create a relationship. Healthcare and financial providers who exhibit empathy towards their clients are more likely to have improved outcomes as well as long-term relations between the client and the service provider.

It is known that empathy, which can be portrayed by facial expressions of chatbots, is perceived to create increased levels of comfort, leading to a smoother interaction. The best way to represent emotion in a digital agent are gestures and facial expressions (Brave & Nass, 2007). In their study from 2007, Brave and Nass found that a smiling face and open posture indicate more empathy for the conversational partner. They identified that raised cheeks are one of the minimal cues efficient enough for signalling friendliness (Brave & Nass, 2007). Small changes in mouth position may have similar effects.

Additionally, Baxter et al. (2018) say that female chatbots are seen as more empathic. Borau et al. (2021) explain that women are better at understanding concerns and can better provide emotional support. While this may be less important for financial matters, it is inevitable in a medial context. Further, independent of context, men are judged with having less emotional intelligence,

which can even be perceived as harmful (Borau et al., 2021). While compassion is of lower importance for financial matters, a lack can have severe consequences in a medical environment, as patients need long-time support. Therefore, empathy is indispensable, as in a healthcare repeated counselling by the same doctor is the norm.

Based on previous exemplifications about visual cues and friendliness, it can be argued that adding a smile to the chatbots' appearance will boost perceptions of empathy for the female chatbot even more.

H7: A smiling facial expression, compared to a neutral facial expression, will lead to a higher level of empathy.

H8: A female chatbot, compared to a male chatbot, elicits a higher amount of perceived empathy.

H9: In a healthcare context, a smiling face leads to higher perceptions of empathy than a neutral facial expression, while in a financial context, no difference is expected.

H10: In a healthcare context, a female chatbot leads to higher perceptions of empathy than a male chatbot, while in a financial context, no difference is expected.

H11: In a healthcare context, a smiling female chatbot, compared to a neutral male chatbot, will lead to higher perceptions of empathy.

2.2.5 Social presence

Social presence refers to the degree to which individuals feel engaged and connected in a computer-mediated communication environment. Social presence is essential for fostering effective communication, collaboration, and community building in virtual environments. It is helpful to incorporate personality traits such as sociability and warmth (Hess et al., 2009). There is a need for chatbots to be perceived as present to encourage users to engage with them and feel comfortable disclosing sensitive information (Ng et al., 2020). In a study by Ogonowski et al. (2024), they depicted that the perceived social presence displayed through having a live chat box has the greatest influence as it depicts a direct interaction and the focus is on the user themselves. A smiling chatbot increases perceptions of social presence as well (Zogaj et al., 2023).

In a medical context, delivering a higher level of social presence by assigning a more human-like personality or image to healthcare agents, visualized by a smiling facial expression, can enable a sense of human touch and warmth during human-agent interactions (Ng et al., 2020).

H12: A smiling facial expression, compared to a neutral facial expression, will lead to higher social presence.

H13: In a healthcare context, a smiling facial expression, compared to a neutral facial expression, will lead to higher social presence, while in a financial context, no difference is expected.

2.2.6 Competence

By understanding and meeting consumer expectations effectively, chatbots can enhance perceptions

of competence (Jiménez-Barreto et al., 2023). The competence of chatbots plays a significant role in their potential impact. Chatbots can support education and training by providing easily accessible and low-threshold conversational interfaces for users (Følstad et al., 2018). Offering detailed and informative responses can contribute to perceived competence by showcasing the chatbot's knowledge and expertise in the subject matter.

Concerning the visual appearance portraying competence, Fujiwara et al. (2022) postulate than competence is commonly associated with a cold character, so a non-smiling facial expression. This originates from the classic image of the scientist being knowledgeable but without any emotions (Fujiwara et al., 2022). Hence, a neutral facial expression is expected to create a higher sense of competence. Nevertheless, this contradicts with the conceptions of how a healthcare worker should behave. Therefore, it is supposable that the cold expression conveying competence may be true for a financial advisor, however, it is unlikely that the effect will be the same in the medical context.

Regarding gender, it is plausible that competence will be perceived higher for the male than for the female chatbot, as the mechanisms of competence work identically to those attributing other characteristics. As explained earlier, some traits, especially desirable ones, are primarily attributed to men (Amin et al., 2014). Based on stereotypes and prejudice, it is often the case that men are referred to be more competent.

> H14: A male chatbot, compared to a female chatbot, will lead to higher perceptions of competence.
> H15: A neutral facial expression, compared to a smiling facial expression, will lead to higher perceptions of competence.

2.2.7 Intention to use

The quality of a chatbot can also be assessed by the users' intention to use the bot for future interactions. Participants' intention to use is a function of their perception that the chatbot could assist them in their goal attainment (Terblanche & Kidd, 2022). Ensuring that the chatbot interface is user-friendly, intuitive, and easy to navigate can increase users' intention to use the technology. Simplifying interactions and providing clear instructions can make the experience more seamless (Schillaci, 2024). It needs to be made sure that the chatbot offers meaningful and valuable interactions to users. By meeting users' needs, providing relevant information, and offering helpful assistance, chatbots can demonstrate their utility and usefulness, thereby increasing users' intention to engage with them.

However, the relevance or severity of a topic can hamper the intention to use a chatbot (Kuhail et al., 2022). Although, this may be reluctance in general, not necessarily caused by the chatbot. Intention to use is reached by a balanced relationship between trust and risk (Kasilingam, 2020). Personalised interactions and emotional bond determine intention to use as well (Schillaci, 2024). If a previous experience has been positive and perceived ease of use is high, the intention to use will rise as well.

The design or persona of a chatbot can impact users' perceptions of how well the bot aligns

with their preferences and needs (Schillaci et al., 2024). A well-designed interface can convey a sense of quality and sophistication, increasing users' intention to use the chatbot to access its services or information (Schillaci et al., 2024). Visually appealing design encourages continued engagement with the service it offers. Perceived chatbot human-likeness positively affects perceived self-congruence between consumers and chatbots, hence, self-congruence mediates the effect of perceived chatbot animation on repeated intention to use (Zogaj et al., 2023).

A study by Alt et al. (2021) identified several factors that influence consumers' intention to use banking chatbot technology. In their study on technology in the banking industry, they found that consumers' perception of the chatbot being compatible with their lifestyle had a strong effect on their willingness to adopt the technology. The intention to use can be increased if using the chatbot leads to advantages and benefits in managing finances (Alt et al., 2021). In the context of healthcare services, users' perceptions of how well a chatbot's gender aligns with its designated role (e.g., prevention, diagnosis, therapy) can influence their intention to use the tool.

In a study by Terblanche & Kidd (2022), where adoption factors for a coaching chatbot have been analysed, the intention to use was higher for a female chatbot tested in a coaching context which can be compared to counselling in a banking context. Likewise, in a study by Pawlik (2021), the female chatbot design had a positive impact on intention to use, independently of context.

H16: A female chatbot, compared to a male chatbot, will lead to higher intentions to use.

H17: In a banking context, a female chatbot, compared to a male chatbot, will lead to higher intention to use, while in a healthcare context, no difference is expected.

All expectations and hypotheses are visualised in Figure 1.

Figure 1
Research model
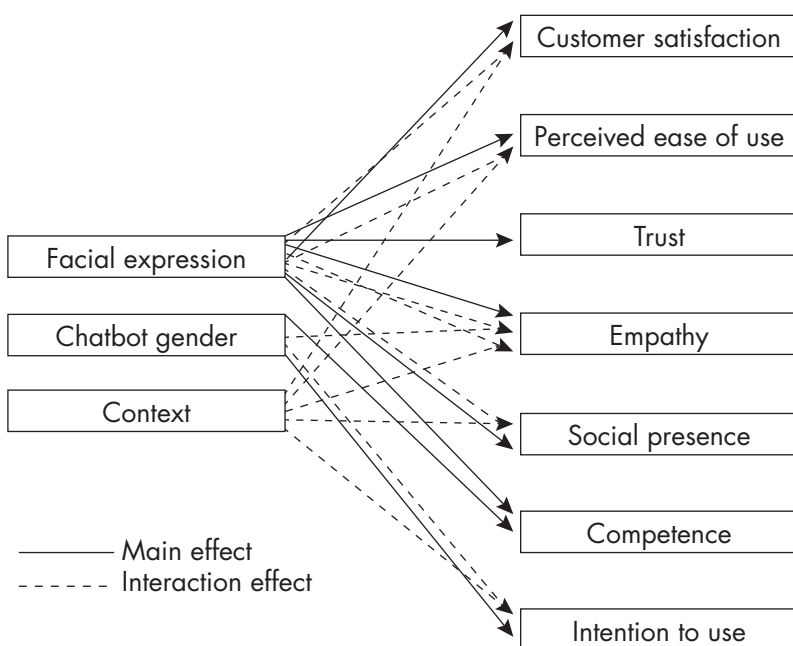
3. Methods

In this research, a quantitative form of data collection was used. A questionnaire with 40 questions was designed to elaborate users' trust and user experience with a chatbot.

3.1 Design

This study employs a 2x2x2 design to examine the effects of visual appearance, context and gender of a chatbot in a cross-sectional study. Visual appearance is compared on the two levels of smiling and neutral, gender is compared male versus female, and context includes the levels of healthcare versus finances. This allows for measuring main and interaction effects on the dependent variables trust, perceived ease of use, empathy, customer satisfaction, competence, social presence, and intention to use. A self-administered questionnaire was utilized as the primary instrument for data collection. Participants were randomly assigned to evaluate one of the chatbots varying in context, gender, and facial expression.

3.2 Participants and recruitment

Participants were recruited by distributing the questionnaire via WhatsApp, email, a post on LinkedIn, and lastly, through snowballing by asking recorded participants to share the link. Further, the questionnaire was made available in the SONA system of the University of Twente. As these methods did not gather enough responses, the survey was also shared on SurveyCircle, a platform for supporting each other and collecting participants by filling out their surveys.

The participants in this study are all people who have access to the internet and are able to use a chatbot. Participants are mainly students, due to the distribution and sharing of the survey. A total of 280 participants was reached, of which 48 participants were excluded, so the final sample consists of 232 responses. 42 participants were excluded because they did not finish the complete survey, four participants were excluded because they answered every question on the same number on the scale and two more participants were excluded because they gave the wrong answers to the manipulation check. In the final sample, 89 participants were male, 139 female, one identified as non-binary and two preferred not to answer. 189 of the participants were German and 43 from another country.

The questionnaire was distributed to participants electronically including a small introduction to the topic and an explanation of the study. Participants were asked to indicate a few demographics and finally complete the questionnaire with all questions. The survey ended up with 232 useful responses. The frequency table showed that all tested conditions have at least 24 responses, one condition has 35 responses. The survey was completed by more female than male participants and above 80% of respondents are German. Most participants indicated that they do have experience with chatbots, while many do even use them frequently. Therefore, it can be assumed that the participants of this study are eligible to evaluate the presented avatars.

Looking at the demographics overview, it became obvious that there was an unequal division of the participants' gender among for some conditions. Following, a chi-square test was carried

out. As the test showed a significant value of 0.03, t-tests for all constructs were conducted. Since only one t-test was significant, and the means of the two groups differed by just 0.2, gender of the participant was not included as a covariate.

Table 1

Demographic characteristics of the Participants (N = 232)

Sample characteristic	Category	Condition								
		1	2	3	4	5	6	7	8	Total
Gender	Male	18	15	14	17	5	8	4	6	87
	Female	14	15	10	7	21	18	30	24	139
	Other	0	0	0	0	2	0	1	0	3
										229
Nationality	German	28	30	17	22	20	19	28	24	188
	Other	4	0	7	2	8	7	7	6	41
										229

3.3 Materials

3.3.1 Pre-test

The pre-test is a qualitative way of data collection by observing the participants' reaction to the avatars as well as the videos and by asking a few interview questions to better understand and evaluate participants' preferences. The pre-test is meant to validate the design ideas and gather feedback and suggestions for improvement. The slides for the pre-test can be found in Appendix A.

3.3.1.1 Method

In this pre-test, a small number of participants were asked to evaluate the design of the chatbot and examine whether the chatbot fits its purpose. The avatars have been designed in Adobe Illustrator. The aim of this pre-test was to test the independent variables and examine whether the differences in design were clear. By this, the strengths and weaknesses of every single chatbot were analysed and the best fits for the main experiment were chosen. A diverse group of participants representing the target audience were recruited. Four of the interviews were conducted in person and two via video. Two participants were male and four were female, 2 were Dutch and four were German, and 5 people under the age of 30 were asked and one person older than 50. Contrary to the ultimate experiments where the dependent variables such as the customer satisfaction and participants' trust in the chatbot were evaluated, the pre-test was investigating the clarity of the design and participants were shown many different versions of the chatbot. The graphics of the different variations of the chatbot designs were compared. Slides with a pre-selection were shown ordered by context, gender and facial appearance. Participants were asked to rank the different designs and ultimately pick a favourite plus make a ranking of the four versions shown per slide. All avatars only differed slightly, and these differences are the same among all avatars on all slides. The differences are the face with

or without glasses, with or without a tie, with or without a stethoscope, and with suit and tie/blouse or a shirt.

An overview of all avatars tested is presented below.

Figure 2
Overview of all tested avatars

Moreover, a video was presented, once concerning a medical context, once a financial context. The purpose of this video was to present the fictional conversation of a user with the bot to the participants. This was the conversation that was eventually used in the main test. Both the conversation in the medical and in the financial context were dealing with an issue comparable in relevance and no technical terms were used in either of the videos. It was made sure as well that the length of the videos only differed by a few seconds.

Regarding the video, it was tested whether there is enough input to understand the conversation. Additionally, participants were asked to evaluate whether the conversation is realistic and if the terms are understandable. Further, the speed of the video was supposed to be evaluated. The video should not be too fast in order to be able to carefully read the conversation but also not too slow to make sure it does not become boring or participants attention decreases. The video was shown once and then participants were asked for feedback. If the video speed was too fast or too slow, another version with the video speed in 0.8x speed or 1.2x speed was presented.

3.3.1.2 Results
The results of the pre-test support the design ideas and feedback from the participants reassures that the differences in the design are clear and the participants can imagine these avatars to work well in the desired context. The goal of the pre-test was to find out whether participants prefer a more formal version with a suit and tie and if wearing glasses does have an influence on the liking of the avatar. Out of the several options, one avatar per condition was selected as the best fit design

wise. The criteria for this were the rating of the participants. The majority of participants preferred the chatbot wearing glasses, hence, an avatar with glasses was selected for all conditions. In the medical context, all participants liked the avatar wearing a stethoscope and in the financial context, the participants preferred the more formal version of the male avatar wearing a tie. Therefore, the more formal version of the female avatar wearing a blazer was selected as well. For the presentation of the avatar in the survey, it was made sure that all chatbots throughout all conditions match exactly in level of formality. Eventually, one version of the male avatar and one version of the female avatar were chosen per context and this version was used in the experiment once as smiling representation and once as a neutral representation of their gender. The selected avatars are presented in the figure below.

Figure 3
Highest ranked avatars

Further, some feedback regarding the conversation shown in the video was implemented and minor parts of the formulation of the medical context were changed or adapted, while the conversation for the financial condition was revised in several sentences. For the medical condition, this means that in one sentence, the word "fine" has been replaced by "nice", as the word fine was considered to be passively offensive. Further, the goodbye of the healthcare chatbot was too short and was altered to be more empathic. In the financial conversation, the answers of the chatbot were evaluated to be a bit hasty. Therefore, some more questions were included to make the conversation between the user and the chatbot smoother.

Regarding the demographics of the participants, no difference between male and female participants was observed, no difference in the age of participants was observed, however, a tendency towards liking more or less professional appearances became obvious between Dutch and German participants. However, the number of participants for the pre-test is far too small for any inferences. Nevertheless, nationality was included as a further background question in the survey.

A summary of every pre-test can be found in Appendix B.

3.4 Measures
To collect the participants' opinion on the different chatbot designs, a questionnaire has been developed.

3.4.1 Questionnaire development and measurement scales

The questionnaire used in this study was developed based on a thorough review of relevant literature and existing validated instruments in the field. The questionnaire consists of items derived from the human robot interaction (HRI) scale, the post study system usability questionnaire

(PSUUQ), and the healthcare relationship (HCR) trust scale. The HRI is a subjective tool that was specifically designed to measure human-robot trust and can be expressed as an overall percentage of trust (Schaefer, 2016). The questions about the system and robotics are adapted to the case of the chatbot, since it is a robotic system.

The PSUUQ is an evaluation tool from 1992 used to assess perceived usability of various interfaces and hardware, including educational technology (Vlachogianni & Tselios, 2023). From this questionnaire, four questions were used to measure perceived ease of use. Items are, for example, "It was simple to use this system" and have been adapted by replacing 'system' with 'chatbot'. Other items like "I am satisfied with this system" have been adapted to "I am satisfied with the experience with the chatbot" in order to measure customer satisfaction.

The Health Care Relationship (HCR) Trust Scale is a 15-item scale that measures patients' level of trust in their healthcare provider (Bova et al., 2012). Questions like "I feel comfortable talking to my healthcare provider about my personal issues" were changed to "I felt comfortable talking to the chatbot about a personal matter".

Parts of the human robot interaction (HRI) scale have been used to complete the survey for more than one variable. Examples for this are elements like the bot "communicates clearly" or the bot is "competent".

The final questionnaire consists of 40 questions of which 30 examine the chatbot. The remaining ten questions investigate the participants' demographics as well as the pre-knowledge of the respondents. Responses to the questionnaire items were collected using a 7-point Likert scale. The scale was anchored with the points 1= strongly disagree to 7= strongly agree. A 7-point scale is considered a good compromise between having enough response options to capture nuanced responses and not overwhelming respondents with too many options (Dawes, 2006). Linder (2020) conducted a study on healthcare chatbot in relation to trust and effectiveness and applied a 7-point Likert scale as well.

The complete questionnaire can be found in Appendix C.

3.4.2 Reliability and validity analysis

Next, the KMO test gave a value of .93, which is very good (Kaiser, 1974) and means that the data is suitable for factor analysis. The scree plot showed that there are 5 components with an Eigenvalue higher than 1 and a rotated component matrix was used to determine which items load on the different constructs. The 7 proposed constructs load on 5 different factors. The construct 'perceived ease of use' loaded on one factor alone as well as the constructs 'empathy' and 'social presence'.

The construct 'intention to use' partially loaded on two factors with two items but could clearly by assigned to one of the factors due to much higher loading numbers. The construct 'trust' loads

one the same factor as 'intention to use'. Therefore, a new construct with the name 'commitment' has been created. When users trust a product or service, they believe it will perform reliably and meet their expectations. This belief directly influences their intention to use the product, as they feel confident that it will deliver the desired results without issues (Ejdys, 2020).

Finally, the constructs 'competence' and 'customer satisfaction' load together on one factor and only on this one factor. Hence, these two constructs have been combined. Competence and customer satisfaction are closely connected because a company's or employee's competence directly impacts the quality of service, problem-solving, and overall customer experience. When customers perceive competence in a service provider, they are more likely to trust the organization and feel satisfied with the outcomes (Gritti and Foss, 2010). As competence enhances customer satisfaction by addressing personal needs accurately, customers that feel understood are more likely to have a positive experience. Following, the new construct is named 'quality'.

The rotated component matrix is displayed below.

Table 2

Rotated component matrix

	1 Quality	2 Commit-ment	3 Empathy	4 Perceived ease of use	5 Social presence
4.1 I would be satisfied with the experience with the chatbot.	.58				
4.2 I would be satisfied with the service the chatbot provided.	.73				
4.3 I would recommend using this chatbot to my friends and family.	.52				
4.4 I think the chatbot is adequately supporting the user.	.76				
4.5 I think that the chatbot sufficiently solves the users' concern.	.70				
4.6 I would have a positive feeling using the chatbot.	.46				
5.1 The chatbot is skilled.	.70				

Item			
5.2 The answers of the chatbot are helpful and thought-out.	.79		
5.3 The chatbot responds considerately to the users' questions.	.69		
5.4 The chatbots' performance was qualified.	.80		
1.1 I would feel comfortable talking to the chatbot.		.78	
1.2 I would feel in control over the interaction with the chatbot.		.66	
1.3 I would feel that I as a user can trust the chatbot completely.		.80	
1.4 I would feel that the chatbot is trustworthy.		.78	
7.1 I think using the chatbot would be a good idea.		.70	
7.2 I would use the chatbot if I have any concerns in the future.		.72	
7.3 I think that I would like to use this chatbot frequently.		.69	
7.4 I think that using the chatbot would be beneficial.		.63	
6.1 I would feel emotionally connected to the chatbot.			.78
6.2 I would feel actively engaged in the conversation.			.67
6.3 I would feel a sense of connection with the chatbot.			.82
6.4 I would feel like I was talking to a real person.			.78

2.1 It seems that it was simple to use this chatbot.	.80	
2.2 It seems as if the chatbot was unnecessarily complex.	.58	
2.3 It seems that the chatbot was easy to use.	.85	
2.4 It seems as if I would be able to interact with the chatbot without any efforts.	.80	
3.1 The chatbot understands the emotions of the user well.		.73
3.2 The chatbot was able to correctly respond to the users' emotions.		.74
3.3 The chatbot cared about the users' emotions.		.69
3.4 The chatbot was able to make the user feel comfortable.		.59

Factor loadings smaller than .4 have been deleted. This selection has been made due to the sample size between 200 and 250 participants.

A reliability analysis was conducted computing the Cronbach's Alpha of the constructs. Therefore, the negative variable ("It seems as if the chatbot was unnecessarily complex") first had to be reverse coded. All constructs have a Cronbach's alpha above .8 meaning there are high levels of internal consistency (Tavakol & Dennick, 2011).

Table 3

Cronbach's alpha table

Construct	Cronbach's Alpha
Trust	.93
Perceived ease of use	.81
Empathy	.84
Satisfaction	.95
Social presence	.86

3.5 Data analysis

Data analysis was conducted using SPSS. Descriptive statistics, such as means, standard deviations, and frequencies were calculated to summarise participant characteristics and responses to

questionnaire items. Inferential statistics, including ANOVA, t-tests and regression were employed to examine relationships and differences between variables.

4. Results

Analysis was conducted with visual appearance, gender, and context functioning as the independent variables and commitment, perceived ease of use, social presence, empathy, and quality as dependent variables. To investigate the differences between the groups and determine whether these differences are statistically significant, five Analyses of Variance (ANOVAs) were conducted.

4.1 Quality

Table 4

Descriptive Statistics for Quality

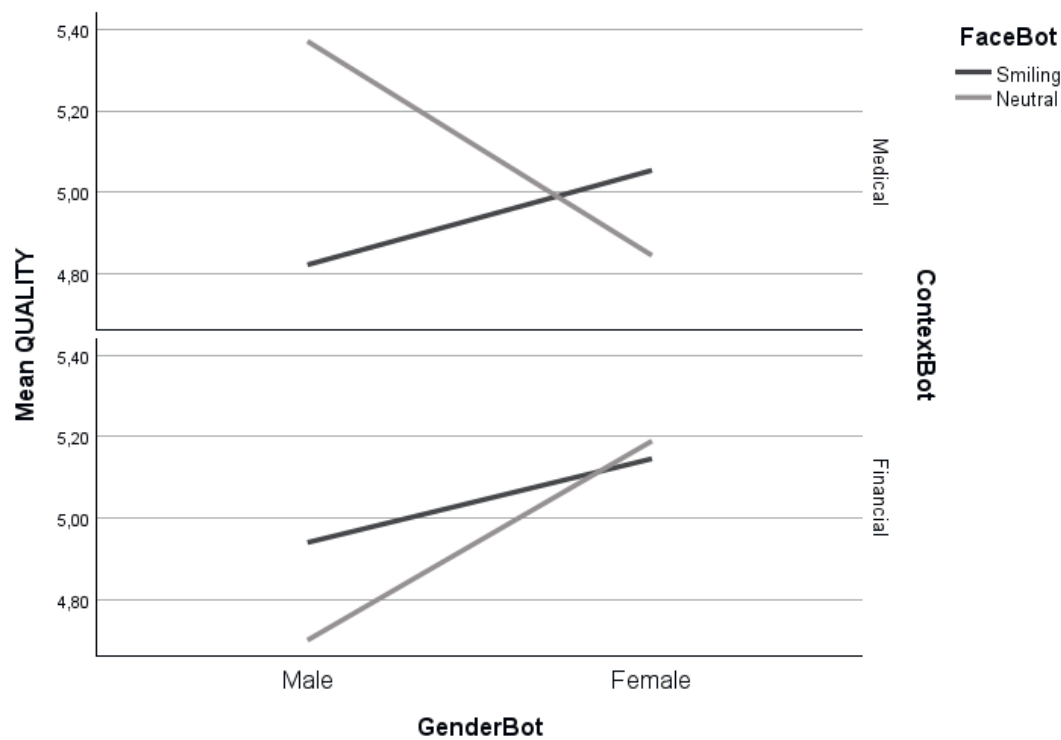| Gender	Visual appearance	Context	M	SD
Male	Smiling	Medical	4.82	1.35
		Financial	4.94	1.10
	Neutral	Medical	5.37	0.91
		Financial	4.70	1.10
Female	Smiling	Medical	5.05	0.95
		Financial	5.15	0.94
	Neutral	Medical	4.84	1.13
		Financial	5.19	0.99

Measured on a 7-point Likert scale, 7= totally agree, 1= totally disagree

Univariate analysis of variance (ANOVA) was conducted with visual appearance, gender, and context as independent variables and quality as dependent variable. It was expected that there would be a main effect of visual appearance, a main effect of gender, an interaction effect with visual appearance in the healthcare context, and an interaction of gender and context.

The main effect of visual appearance on quality was non-significant, ($F(1, 220) = .06$, $p = .80$). The main effect of gender on quality was also non-significant ($F(1, 220) = .49$, $p = .49$). There was also no interaction effect between visual appearance and context ($F(1, 220) = .88$, $p = .35$), but there was a tendency towards a significant interaction effect of context and gender ($F(1, 220) = 2.98$, $p = .09$, $\eta^2 = .01$). The numbers indicate that the quality of the female chatbot is evaluated higher than the male one in the financial context. Moreover, the three-way interaction of all independent variables, gender, context, and visual appearance showed a tendency as well ($F(1, 220) = 3.30$, $p = .07$, $\eta^2 = .02$). However, this effect is contrary to what was hypothesized. The smiling female chatbot is evaluated higher than the smiling male chatbot in the medical as well as in the financial context and the neutral female chatbot is also evaluated higher than the neutral male chatbot in the financial chatbot. Only in the medical context with a neutral visual appearance, the male chatbot was evaluated higher than the female chatbot. This last effect was expected to be the other way around. The interaction is visualised in Figure 4. The results mean that H1, H2, H3, H14, and H15 all must be rejected.

Figure 4

Three-way interaction of visual appearance, gender and context with quality

4.2 Perceived ease of use

Table 5

Descriptive Statistics for Perceived ease of use

Gender	Visual appearance	Context	M	SD
Male	Smiling	Medical	4.24	1.57
		Financial	4.41	1.11
	Neutral	Medical	4.88	1.10
		Financial	4.26	1.40
Female	Smiling	Medical	4.64	1.42
		Financial	4.51	1.36
	Neutral	Medical	4.60	1.40
		Financial	4.60	1.19

Measured on a 7-point Likert scale, 7= totally agree, 1= totally disagree

For perceived ease of use, a main effect of visual appearance as well as an interaction of gender and context were expected.

There was no main effect of visual appearance on perceived ease of use ($F(1, 220) = .55$, $p = .46$), as well as no interaction effect between context and gender ($F(1, 220) = .22$, $p = .64$). Therefore, H4 and H5 must be rejected. This means that there are no meaningful differences

between perceptions regarding any of the independent variables.
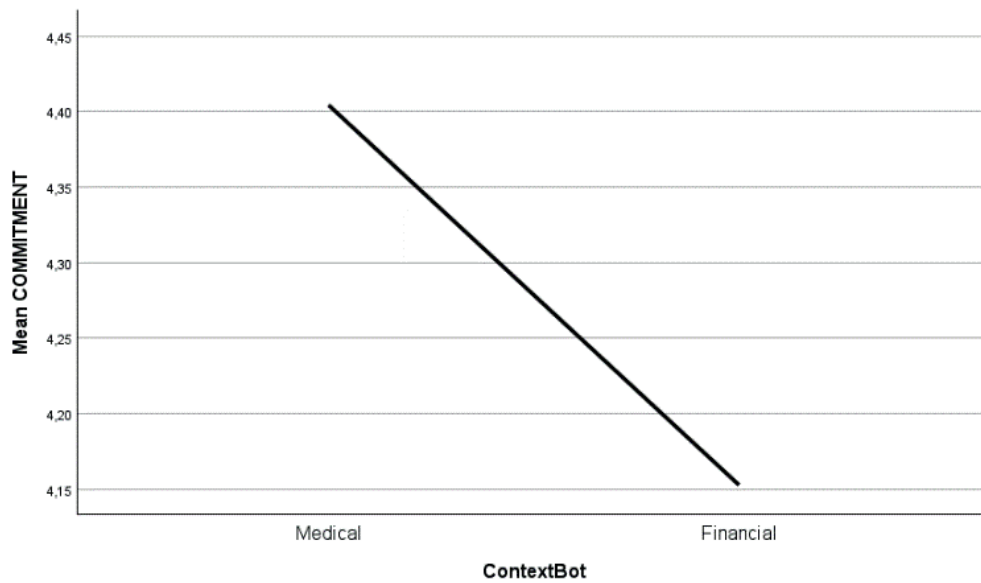
4.3 Commitment

Table 6
Descriptive Statistics for Commitment

Gender	Visual appearance	Context	M	SD
Male	Smiling	Medical	4.19	1.41
		Financial	4.01	1.12
	Neutral	Medical	4.66	1.15
		Financial	4.09	1.31
Female	Smiling	Medical	4.54	1.29
		Financial	4.16	1.18
	Neutral	Medical	4.31	1.23
		Financial	4.33	1.23

Measured on a 7-point Likert scale, 7= totally agree, 1= totally disagree

For commitment, a main effect of gender was expected, a main effect of visual appearance, an interaction effect of visual appearance and context, and an interaction effect of gender with the banking context.

There was no main effect of visual appearance ($F(1, 220) = .59$, $p = .45$), the main effect of gender on commitment was non-significant ($F(1, 220) = .36$, $p = .55$) as well, however, there was a tendency of an effect of context ($F(1, 220) = 2.66$, $p = .10$, $\eta^2 = .01$). The results indicate that commitment was perceived higher in the medical context, regardless of gender and visual appearance. This is visualised in Figure 5. There is no interaction effect between context and visual appearance ($F(1, 220) < .01$, $p = .96$), as well as no interaction effect of context and gender ($F(1, 220) = .39$, $p = .53$). The results are not in line with H6, H16, and H17 and the hypotheses must be rejected.

Figure 5

Main effect of context on commitment

4.4 Empathy

Table 7

Descriptive Statistics for Empathy

Gender	Visual appearance	Context	M	SD
Male	Smiling	Medical	4.68	1.36
		Financial	4.27	1.14
	Neutral	Medical	4.77	1.22
		Financial	3.71	1.07
Female	Smiling	Medical	4.83	0.91
		Financial	4.48	1.19
	Neutral	Medical	4.38	1.35
		Financial	4.44	1.18

Measured on a 7-point Likert scale, 7= totally agree, 1= totally disagree

The variable empathy was expected to have a main effect of visual appearance, a main effect of gender, an interaction effect of visual appearance within the healthcare context, and an interaction of gender within the healthcare context.

There was no main effect of visual appearance on empathy ($F(1, 220) = 2.21$, $p = .14$) and the main effect of gender on empathy was non-significant ($F(1, 220) = 1.22$, $p = .27$) The main effect of context was significant ($F(1, 220) = 7.58$, $p = .01$, $\eta^2 = .03$). The effect of empathy is higher in the healthcare context. There was also no interaction effect of visual appearance and context ($F(1, 220) = .15$, $p = .70$), but there was a tendency towards a significant interaction effect of context and gender ($F(1, 220) = 3.40$, $p = .07$, $\eta^2 = .02$). The numbers show that empathy is

perceived higher for a female chatbot in the healthcare context. The numbers here are higher, both for the neutral as well the smiling female chatbot. There was also a tendency towards a significant interaction effect of all independent variables, visual appearance, gender, and context ($F(1, 220) = 2.78$, $p = .09$, $\eta^2 = .01$). For the interaction, although there is only a tendency towards an effect, the numbers for empathy for a smiling female chatbot in healthcare are meaningfully higher than for a neutral male chatbot in finance. These effects have been expected. The results mean that H7, H8, and H9 must be rejected but H10 and H11can be accepted. These hypotheses suggested that, in a healthcare context, a female chatbot will lead to higher perceptions of empathy as well as a smiling chatbot would lead to higher perceptions of empathy in the healthcare context as well. These effects are visualised below.

Figure 6

Interaction of visual appearance, gender and context with empathy
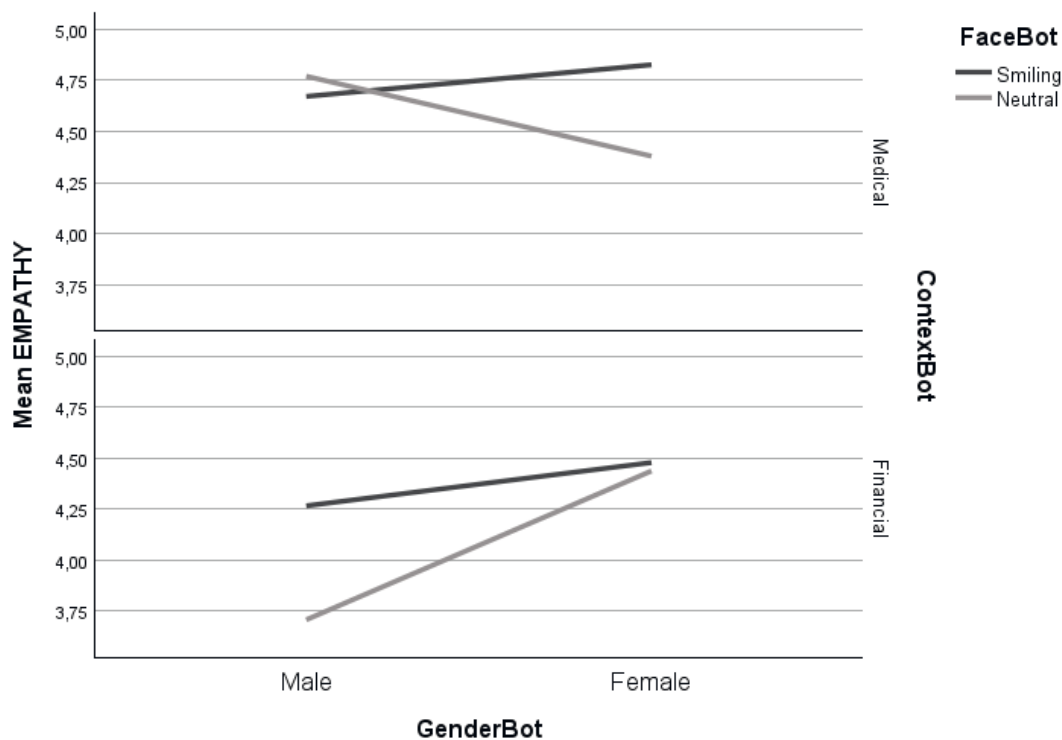
4.5 Social presence

Table 8

Descriptive Statistics for Social presence

Gender	Visual appearance	Context	M	SD
Male	Smiling	Medical	3.23	1.27
		Financial	3.02	1.28
	Neutral	Medical	3.71	1.42
		Financial	2.71	1.28

Female	Smiling	Medical	3.44	1.57
		Financial	2.92	1.18
	Neutral	Medical	3.11	1.34
		Financial	3.50	1.50

Measured on a 7-point Likert scale, 7= totally agree, 1= totally disagree

Regarding social presence, it was expected that there would be a main effect of visual appearance as well as an interaction effect of visual appearance and context.

There was no significant main effect of visual appearance on social presence ($F(1) = .27$, $p = .61$) but there was a tendency towards a main effect of context ($F(1) = 3.69$, $p = .05$, $\eta^2 = .02$). There was no significant interaction effect between visual appearance and context ($F(1) = .01$, $p = .93$) but a significant interaction effect of visual appearance, gender, and context ($F(1) = 5.26$, $p = .02$, $\eta^2 = .02$). The interaction is visualised in Figure 7.

According to the numbers, except for the neutral, female chatbot in the financial context, social presence was evaluated higher for the healthcare chatbot in the remaining three conditions. In the financial context, social presence is higher for the neutral female avatar than for the neutral male avatar. On the other hand, social presence is higher for the neutral male avatar in the healthcare context. Social presence is also higher for the smiling female chatbot than for the smiling male chatbot in healthcare, however, it is also higher for the smiling male chatbot in finance than for the smiling female chatbot in finance. These effects are contrary to the expectations. The results are not in line with H12 and H13 and the hypotheses must be rejected.

Figure 7

Interaction of visual appearance, gender, and context with social presence
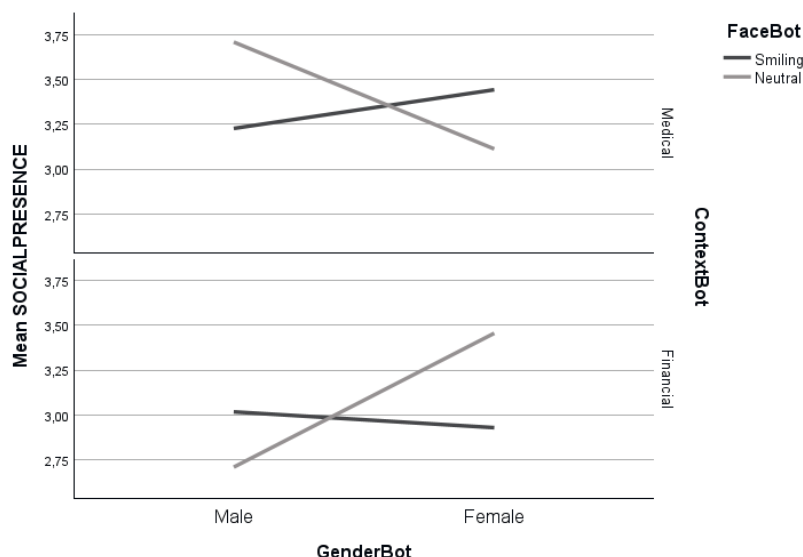
To sum up, the findings showed that there are no main effects of the independent variables visual appearance and gender. There are also no interaction effects of visual appearance with a second

independent variable. Visual appearance was only significant in the three-way interaction with all independent variables, and this also only in relation to quality and social presence. This means that visual appearance only had a relevant effect in combination with the other two independent variables gender and context. In relation to quality, the smiling female chatbot in healthcare produced higher numbers than the neutral male chatbot in finance and with regard to social presence, the numbers differed in any combination of variables. Since the variable does not produce significant effects on its own, it means that, across all levels of the other variables, it does not have a uniform impact on the dependent variable. Following, visual appearance did not produce any main effects but does have significant interaction effects. Its influence is only apparent under certain conditions, specifically when interacting with particular combinations of the other variables.

The independent variable gender only produced significant effects in combination with the variable context, and this in the two-way as well as in the three-way interaction. One gender usually produced higher numbers than the other in the different contexts with relation to the dependent variables. With regard to empathy, the numbers were the highest for the smiling female chatbot in healthcare. This is the only three-way interaction where gender did have an effect.

The independent variable context did have the most effects, context affects all dependent variables except for perceived ease of use. Here, the healthcare context generates higher scores in any relationship, except for the relation with quality. For the quality of the avatars, three of four times the avatar in the financial context scored higher. However, the score of the male neutral avatar in the healthcare context is the highest of all scores relating to quality.

The most effects on a dependent variable can be observed for empathy, where context did have a significant main effect, an interaction effect with gender and a significant three-way interaction with all independent variables. Empathy is evaluated higher for the female chatbot and the healthcare chatbot has evoked higher ratings of empathy than the financial chatbot. Empathy is the highest for a smiling female chatbot in medicine, and this confirms one of the expectations. For social presence, the most differences have been found. In general, the measures in the healthcare context are higher as well, and the smiling avatar, as well as the female avatar scores better, except for the neutral male avatar in the healthcare context. Surprisingly, this avatar does have the highest score of all regarding social presence.

Interestingly, there are no effects of any independent variable on the dependent variable perceived ease of use. There are no results for any of the hypotheses as well as no unexpected results. According to the results of this study, neither visual appearance nor context nor gender do influence perceptions of perceived ease of use in any way.

Overall, the data support the notion that at least one of the independent variables, context, does influence perceptions of a chatbot's quality, commitment, empathy and social presence. Moreover, empathy, social presence, and quality of the chatbot have a larger influence than perceived ease of use and commitment. These findings provide a foundation for understanding the role of human traits in the application of a digital assistant.

An overview of all hypotheses and which of those have been confirmed as well as which had to be rejected is presented below in Table 9. The effects found are visualised in Figure 8.

Table 9

Overview of hypotheses

No	Hypothesis	Outcome
H_1	A smiling chatbot, compared to a neutral chatbot, will lead to a better customer satisfaction.	Rejected
H_2	In a financial context, a male banking chatbot will lead to a higher customer satisfaction, while in a healthcare context, a female healthcare chatbot will lead to a higher customer satisfaction.	Rejected
H_3	In a healthcare context, a smiling female chatbot, compared to a neutral male chatbot, will lead to a higher customer satisfaction.	Rejected
H_4	A smiling facial expression, compared to a neutral facial expression, will lead to higher perceived ease of use.	Rejected
H_5	In a healthcare context, a female chatbot will lead to higher perceived ease of use, while in a financial context, a male chatbot will lead to higher perceived ease of use.	Rejected
H_6	A smiling facial expression, compared to a neutral facial expression, will lead to more trust.	Rejected
H_7	A smiling facial expression, compared to a neutral facial expression, will lead to a higher level of empathy.	Rejected
H_8	A female chatbot, compared to a male chatbot, elicits a higher amount of perceived empathy.	Rejected
H_9	In a healthcare context, a smiling face leads to higher perceptions of empathy than a neutral facial expression, while in a financial context, no difference is expected.	Rejected
H_{10}	In a healthcare context, a female chatbot leads to higher perceptions of empathy than a male chatbot, while in a financial context, no difference is expected.	Confirmed
H_{11}	In a healthcare context, a smiling female chatbot, compared to a neutral male chatbot, will lead to higher perceptions of empathy.	Confirmed
H_{12}	A smiling facial expression, compared to a neutral facial expression, will lead to higher social presence.	Rejected
H_{13}	In a healthcare context, a smiling facial expression, compared to a neutral facial expression, will lead to higher social presence, while in a financial context, no difference is expected.	Rejected
H_{14}	A male chatbot, compared to a female chatbot, will lead to higher perceptions of competence.	Rejected

H_{15}	A neutral facial expression, compared to a smiling facial expression, will lead to higher perceptions of competence.	Rejected
H_{16}	A female chatbot, compared to a male chatbot, will lead to higher intentions to use.	Rejected
H_{17}	In a banking context, a female chatbot, compared to a male chatbot, will lead to higher intention to use, while in a healthcare context, no difference is expected.	Rejected

Figure 8

Main and interaction effects
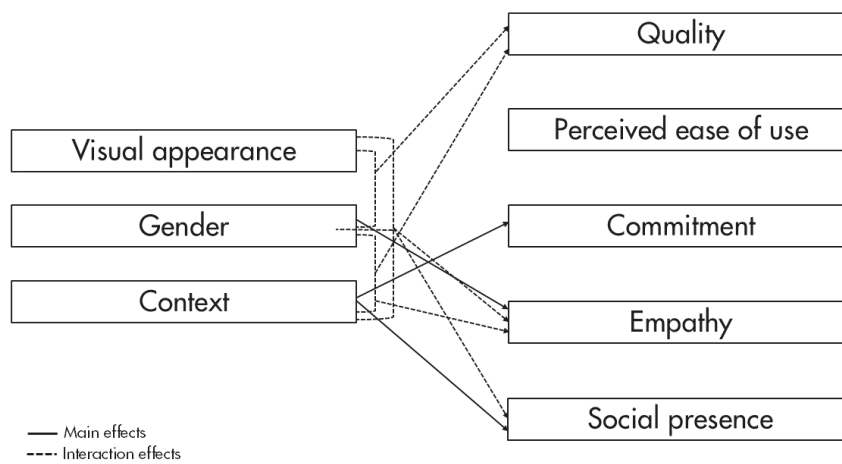
5. Discussion

This study aimed to investigate whether there are differences in chatbot perceptions based on gender of the chatbot, visual appearance of the chatbot, and the context in which the chatbot operates. Moreover, the aim was to find out what the effects of this potential differences are.

It was expected that, in general, the smiling chatbot would be better evaluated in the healthcare context and a neutral chatbot would be better fitted for the financial context. Moreover, it was hypothesized that empathy would be rated higher for the female chatbot compared to the male chatbot. Further, interactions of the independent variables were expected for all dependent variables. It was hypothesized that a female chatbot would be preferred in a healthcare context and a male chatbot would be preferred in a financial context. Moreover, it was expected that smiling would be important in a healthcare context and a neutral facial expression would be a better fit for a chatbot in a financial context. Empathy was expected to be especially important in a healthcare context, while competence was expected to be crucial in a financial context.

5.1 Discussion of results

The main finding of this study is that context does have the largest impact on the evaluations of the chatbot. There is an effect of context in every interaction. Two out of the 19 hypotheses, hypotheses 10 and 11, could be confirmed. H10 hypothesised a female chatbot to lead to higher perceptions of empathy in a healthcare context, while H11 hypothesised a smiling female chatbot to lead to higher perceptions of empathy in a healthcare context. In both of these hypotheses, an effect of context was expected, and both of these hypotheses include empathy as a variable. Empathy is the dependent variable where most effects have been hypothesised and empathy is also the dependent variable that has been affected the most by the independent variables

The healthcare context scores higher in all relations. It is possible that the healthcare conversation has been perceived more realistic. Since the conversations are not taken from a real chatbot but fictitious and written by the researcher, it is possible that the content of the two conversations is not directly comparable. The context alone as well as in interaction with visual appearance or gender did influence the effectiveness of the chatbot the most. Visual appearance did not have any main effects and was only significant in the three- way interaction and only in relation to quality, empathy and social presence. Contrary to the expectations, no meaningful differences were found in users' perceptions of the visual appearance between the two versions. It seems that aspects like appearance and gender, that contribute to the look of the chatbot, are not sufficient to alter perceptions of the chatbot's proficiency. Additionally, gender did not have any main effects, there are only interactions effects where gender is being influential. Following, gender may not have an effect, but participants did notice a difference. One possible explanation for this finding is that it may be that, if the usability of the chatbot is good and the user feels understood and supported, it does not matter if this help has been provided by a male or female persona. Gender was only significant in the interaction with context, two-way as well as three-way. Gender, like visual appearance, did only have an effect on quality, empathy, and social presence. Moreover, these

three dependent variables, quality, empathy, and social presence are those, where a three-way interaction effect with all independent variables was significant or, at least, showed a tendency. This suggests that these dependent variables may share some underlying constructs. Several previous studies (e.g., Belousov, 2023) predicted a relationship between empathy and social presence. Cummings et al. (2021) have specified this relationship to exist with regard to virtual environments and suggest that, when users feel a sense of being together with others, it enhances their ability to understand and emotionally resonate. A similar study (Pimentel et al., 2021) has pointed out social presence as a predictor of empathy. In this research, visual cues have also been put in close relation to these perceptions. They explain that, when a robot exhibits empathic behaviours, such as understanding and responding to a user's emotional states, it can create a stronger sense of social presence (Pimentel et al., 2021). Users are more likely to feel that the robot is a genuine social partner rather than just a machine. Leite et al. (2014) further argue, social presence and empathy add to the perceived authenticity of the chatbot and this transfers to higher perceptions of the chatbot's quality.

While most of the hypotheses of this research could not be confirmed, the unexpected effects give interesting insights into users' perceptions and favourabilities when it comes to chatbot usage. It is surprising that the quality of the male chatbot with a neutral visual appearance was evaluated better than the female chatbot with a smiling visual appearance in the healthcare context. A study (Schillaci et al., 2024) suggests that a chatbot with female characteristics is suitable for prevention roles, while a male chatbot is better suitable for a therapy role. Since the conversation in the study was presented as the user addressing the healthcare chatbot with a question about a personal issue, it can be assumed that the participants in this study experienced the interaction as therapeutic. This means that the participants of this study believed the scenario presented in this study was a patient asking for actual treatment, whereas the scenario in the financial context simply involved a client asking for advice. Further, a chatbot that is not smiling can be perceived as having higher authenticity (Zhang et al., 2024), as people have experienced this behaviour with their general practitioners in the real world. Personal experiences can influence how people subjectively perceive smiling as adequate.

As Skujive et al. (2021) explained, users wish that, in case of confronting the chatbot with a current problem, the interaction leads to initial wellbeing. If this is successful, the commitment to the chatbot increases. In a study by Nadeem et al. (2020), it was explained that social presence influences commitment. This is in line with the finding of this study than both commitment and social presence reached higher numbers in the same context, this context being healthcare.

The findings of this research, although not all significant, are comparable to other studies. In a study by Nißen et al., there was also no effect of gender although anticipated. Regarding the visual appearance of chatbots, a study by Bortko et al. (2023) suggests that users only become aware of the facial expressions when they see it change right before their eyes. This explains why the difference in perceptions between the smiling and the neutral chatbot are not meaningful.

The positive association between empathy and a smiling female avatar align with prior

research. Females appear warmer and this warmth leads to users believing that their needs are better met (Helme-Guizon et al., 2024). On the other hand, the male neutral character in the healthcare context did reach better number for quality. Fink et al. (2020) suggest that the association may be that men should work as doctors as they are more competent. Therefore, the male neutral doctor is congruent with participants' understanding of a good doctor, as it confirms their expectations.

The outcome that most hypotheses are not confirmed may point to the complexity of the phenomena being studied. Most studies analysed in the theoretical framework have been conducted in the USA and some in Asia. It is possible that the dynamics of the variables differ in another cultural context. India is leading worldwide in chatbot usage and citizens in the US are also using chatbots on many occasions (Carbonaro, 2024), while they are less popular in Europe. In Germany, chatbot usage is still sparse and those that are used are mainly for customer service

purposes (Büchel & Engler, 2024). In a more established environment, humans may regard chatbots as other humans, while in this study, participants seem to grant less importance to appearance and gender, proposing that human attributes do not play a role in the interaction, as the chatbot is regarded as non-human.

5.2 Theoretical implications

The findings that highlight the role of context, gender, and visual appearance in shaping user interactions with chatbots have several important ramifications across various fields, including human-computer interaction, computer-mediated communication and social psychology.

Social presence theory refers to the degree to which a communication medium allows users to perceive the presence of others in a mediated environment (Lowenthal, 2009). It encompasses the ability of a medium to convey social cues that facilitate interpersonal interactions, making users feel as though they are engaging with real people rather than just technology (Nadeem et al, 2020). In the context of chatbots, social presence involves creating a sense that users are interacting with a human-like entity, even though they are aware it is an automated system. Considering whether or not a chatbot should smile or show emotions helps increasing perceptions of social presence when the user perceives the behaviour of the chatbot to be adequate. To enhance perceptions of social presence in a chatbot, developers need to include design elements like visual cues, for example, a welcoming and realistic appearance, and situational cues, meaning that the chatbot ensures the user that they understand the request and provide tailored feedback. By being more engaging and attentive, the chatbot appears to be socially present.

Further, Expectation-Confirmation Theory (ECT) is a psychological framework that explains how individuals form satisfaction judgments based on their expectations and the performance of a product or service (Baharum & Jaafar, 2015). The expectations are the anticipations or beliefs that consumers have regarding the performance of a product or service before they use it. The performance refers to the actual experience or performance of the product or service after use. Chatbots need to credibly interact with the user and do not give any room to doubt the capabilities

of the chatbot. The user should not notice any difference in behaviour compared to a real human (Nadeem et al, 2020). The results of this study imply expectations have not been confirmed. This is obvious since several effects showed a relationship that was contrary to what has been hypothesized. It is essential to know what users expect. Here, another concept comes into play.

User-centred design can assist with the development of successful chatbots, as it focuses on the users wishes and needs. Context and requirements need to be taken into account for every development of a digital assistant (Böhm & Kohl., 2020). The integration of chatbots into existing workflows and systems can be complex, requiring careful planning and execution to ensure seamless operation. Companies need to conduct thorough user research to understand the target audience. Creating user personas can help designers visualize the needs, goals, and pain points of different user segments, guiding the design process (Janssen et al., 2022).

These implications offer a path for refining existing theories and developing new frameworks that better capture the complexity of user interaction with technology across diverse situations and contexts.

5.3 Practical implications

Using chatbots for personal support is a growing area of interest in the fields of healthcare, counselling, and customer service. Despite these many areas of application, this study has shown that there is no 'one fits all' version for a chatbot. Developers of a chatbot should analyse what is expected in a specific context and what people are used to from being in contact with human consultants. As in the real world, a chatbot should meet some expectations. Although artificial, these digital assistants should portray knowledge and be able to build a relationship with the user.

A chatbot for healthcare purposes is favourable as it bypasses the burden to see an actual practitioner. A first assessment can be gained without obligations (Lai et al., 2021). Healthcare providers can implement chatbots to offer preliminary mental health support, conduct follow-up checks, or serve as an additional resource for patients needing immediate assistance, reducing the burden on human therapists, while it is rather difficult to implement a suitable chatbot in the banking sector that will have a chance to be accepted by users (Qureshi et al., 2024). To create effective and ethical chatbots, designers need to carefully consider the implications of design and aim for approaches that prioritize user needs, inclusivity, and flexibility. In a context like finance, where no immediate action may be required but some advice is asked for, users will focus more on the reliability of the given information. In healthcare, commitment and empathy are important as users may be in a situation where they feel uncomfortable. Here, feeling understood is what matters most. As empathy has been impacted the most by context, gender, and visual appearance, designers should consider including this element in their chatbots, as many factors may influence this factor.

5.4 Limitations and future research

This study had a relatively small sample with less than 250 participants. Moreover, the demographics of the participants are not very diverse, as mostly German people answered the survey and the majority of participants were students. Therefore, the findings are not generalizable

to elderly age groups. Further, the survey was answered by random people and the answers may not account for situations where people actually seek help in that moment and are in need of a personally tailored answer.

Second, the short duration of the chatbot interactions may not capture long-term engagement patterns. An extension of this cross-sectional study to a longitudinal study and a longer interaction time could analyse whether a relationship between the user and the chatbot can evolve. Also, as the conversation has been simulated, users may want to differ in how they talk to the chatbot and ask other questions than those that have been showcased in this study. Lastly, the effect sizes of this research are rather small, meaning that the relationships measured are relatively weak.

Since only an online questionnaire was used to measure the effects of chatbot design, a qualitative study could be conducted to support these quantitative findings and evaluate the reasons for the unexpected effects. Moreover, the effect of the participants' gender could be included as a moderator variable in a further study to evaluate if there is a correlation between the chatbot's gender and the participant's gender. Finally, as most of the effects included the variable empathy, it could be analysed how empathy for a chatbot can be enhanced and why there are so many differences for this variable. Further studies could also explore how cultural differences shape user perceptions of chatbot efficiency. Additionally, studies could investigate if there are meaningful differences between young adults using chatbots and elderly people interacting with a chatbot.

6. Conclusion

This study explored meaningful elements of chatbot design, emphasizing the importance of different design options. By using a quantitative study, the effects of visual appearance, gender, and context on trust, customer satisfaction, competence, perceived ease of use, empathy, and intention to use have been tested. The study has demonstrated the impact of chatbot characteristics on user perceptions. The role of context emerged as a key factor in influencing user perceptions. The results show that context does have the main influence on perceptions of the chatbot design. The numbers of all dependent variables in the healthcare context are higher, while the effects of visual appearance and gender are neglectable. Visual appearance and gender are only noteworthy in the interaction with other variables; hence, they do not have a standalone effect.

Although this study does not show many significant effects, this research contributes to the current field of research on improving digital assistants. By understanding user preferences and needs, designers can create chatbots that not only solve functional problems but also provide meaningful, interactive experiences. However, challenges remain in meeting users' requirement and creating chatbots congruent with users' expectations. As chatbot technology continues to evolve, future research should focus on refining design elements, exploring contextual variations in user interaction, and integrating more advanced capabilities to further enhance user experiences.

Overall, effective chatbot design is crucial for their success across various industries, from finances to healthcare, and continued innovation in this field holds the potential to transform the way humans interact with digital assistants.

References

Abdellatif, A., Costa, D., Badran, K., Abdalkareem, R., & Shihab, E. (2020, June). Challenges in chatbot development: A study of stack overflow posts. In Proceedings of the 17th international conference on mining software repositories (pp. 174-185). https://doi.org/10.1145/3379597.3387472

Amin, M., Rezaei, S., & Abolghasemi, M. (2014). User satisfaction with mobile websites: The impact of perceived usefulness (PU), perceived ease of use (PEOU) and trust. *Nankai Business Review International, 5*(3), 258-274. https://doi.org/10.1108/nbri-01-2014- 0005

Aumüller, A., Winklbauer, A., Schreibmaier, B., Batinic, B., & Mara, M. (2024). Rethinking feminized service bots: user responses to abstract and gender-ambiguous chatbot avatars in a large-scale interaction study. *Personal and Ubiquitous Computing,* 1-12. https://doi.org/10.1007/s00779-024-01830-8

Baharum, A., & Jaafar, A. (2015). User interface design: A study of expectation-confirmation theory. *5th International Conference on Computing and Informatics.* (pp. 17-24). http://repo.uum.edu.my/15486/

Bastiansen, M. H., Kroon, A. C., & Araujo, T. (2022). Female chatbots are helpful, male chatbots are competent? The effects of gender and gendered language on human-machine communication. *Publizistik, 67*(4), 601-623. https://doi.org/10.1007/s11616-022- 00762-8

Baxter, D., McDonnell, M., & McLoughlin, R. (2018). Impact of Chatbot gender on user's stereotypical perception and satisfaction. *Electronic Workshops in Computing.* https://doi.org/10.14236/ewic/hci2018.154.

Belanche, D., Casaló, L. V., Flavián, M., & Loureiro, S. M. C. (2023). Benefit versus risk: a behavioral model for using robo-advisors. *Service Industries Journal,* 1–28. https://doi.org/10.1080/02642069.2023

Beldad, A., De Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in human behavior, 26*(5), 857-869. https//doi.org/10.1016/j.chb.2010.03.013

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2022). The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. *Verhaltenstherapie, 32*(Suppl. 1), 64-76. https://doi.org/10.1159/000501812

Bhakta, R., Savin-Baden, M., & Tombs, G. (2014, June). Sharing secrets with robots?. In *EdMedia+ Innovate Learning* (pp. 2295-2301). Association for the Advancement of Computing in Education (AACE).

Bickmore, T., Fernando, R., Ring, L., & Schulman, D. (2009, May). Towards empathic touch by relational agents. In *Proceedings Autonomous Agents and Multiagent Systems (AAMAS) Workshop on Empathic Agents.*

Böhm, S., Eißer, J., Meurer, S., Linnyk, O., Kohl, J., Locke, H., Novakovskij, L., & Teetz, I. (2020). Intent identification and analysis for user-centered chatbot design: A case study on the example of recruiting chatbots in Germany. In *Proc. CENTRIC* (pp. 34-48).

Bova, C., Route, P. S., Fennie, K., Ettinger, W., Manchester, G. W., & Weinstein, B. (2012). Measuring patient–provider trust in a primary care population: Refinement of the health TT care relationship trust scale. *Research in nursing & health, 35*(4), 397-408. https://doi.org/10.1002/nur.21484

Bortko, K., Fornalczyk, K., Jankowski, J., Sulikowski, P., & Dziedziak, K. (2023). Impact of changes in chatbot's facial expressions on user attention and reaction time. *PLoS ONE, 18*(7), e0288122. https://doi.org/10.1371/journal.pone.0288122

Brachten, F., Kissmer, T., & Stieglitz, S. (2021). The acceptance of chatbots in an enterprise context–A survey study. *International Journal of Information Management, 60*, 102375.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In Internet Science: *4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings 4* (pp. 377-392). Springer International Publishing.

Brave, S., & Nass, C. (2007). Emotion in human-computer interaction. In *The human-computer interaction handbook* (pp. 103-118). CRC Press.

Büchel, J., & Engler, J. (2024). *Generative KI in Deutschland: Künstliche Intelligenz in Gesellschaft und Unternehmen* (No. 23/2024). IW-Report.

Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C., & McTear, M. (2018). Best practices for designing chatbots in Mental Healthcare – a case study on iHELPR. *Electronic Workshops in Computing.* https://doi.org/10.14236/ewic/hci2018.

Carbonaro, G. (2024). OpenAI's ChatGPT chatbot tops the list but these are the 9 other most popular AI tools just now. *Euronews.* Retrieved from https://www.euronews.com/next/2024/02/02/these-are-the-10-most-widely-used-ai- tools-and-the-people-who-using-them-the-most

Cherniak, K. (2022, September). *Chatbot statistics: What businesses need to know about digital assistants.* Master of Code. Retrieved October 14, 2024, from https://masterofcode.com/blog/chatbot-statistics

Chung, S. J., & Lee, H. (2020). Visual presentation of mental healthcare chatbots for user experience. *Journal of the HCI Society of Korea, 15*(2), 39–45. https://doi.org/10.17210/jhsk.2020.06.15.2.39

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International journal of human-computer studies, 58*(6), 737-758. https://doi.org/10.1016/s1071-5819(03)00041-7

Cummings, J. J., Tsay-Vogel, M., Cahill, T. J., & Zhang, L. (2022). Effects of immersive storytelling on affective, cognitive, and associative empathy: The mediating role of presence. *New media & society, 24*(9), 2003-2026. https://doi.org/10.1177/146144482098681

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International journal of market research, 50*(1), 61-104. https://doi.org/10.1177/147078530805000106

Dinh, C. M., & Park, S. (2023). How to increase consumer intention to use Chatbots? An empiri-

cal analysis of hedonic and utilitarian motivations on social presence and the moderating effects of fear across generations. *Electronic Commerce Research,* 1-41. https://doi.org/10.1007/s10660-022-09662-5

Ejdys, J. (2020). Trust-Based determinants of future intention to use technology. *Foresight-Russia, 14*(1), 60–68. https://doi.org/10.17323/2500-2597.2020.1.60.68

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2019). Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design. In *Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5* (pp. 145-156). Springer International Publishing.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender bias in chatbot design. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers 3* (pp. 79-93). Springer International Publishing.

Fink, M., Klein, K., Sayers, K., Valentino, J., Leonardi, C., Bronstone, A., Wiseman, P. M., & Dasa, V. (2020). Objective data reveals gender preferences for patients' primary care physician. *Journal of Primary Care & Community Health, 11.* https://doi.org/10.1177/2150132720967221

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

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gritti, P., & Foss, N. (2010). Customer satisfaction and competencies: an econometric study of an Italian bank. *Applied Economics Letters, 17*(18), 1811-1817. https://doi.org/10.1080/13504850903357335

Hari, H., Iyer, R., & Sampat, B. (2022). Customer brand engagement through chatbots on bank websites–examining the antecedents and consequences. *International Journal of Human–Computer Interaction, 38*(13), 1212-1227. https://doi.org/10.1080/10447318.2021.1988487

Hartikainen, M., & Väänänen, K. (2023). Towards human-centered design of AI service chatbots: defining the building blocks. In *International Conference on Human-Computer Interaction* (pp. 68-87). Springer, Cham.

Helme-Guizon, A., Broyer, J., Bataoui, S., & Hakimi, M. A. (2024, June). He or she? Impact of gender's well-being chatbots on user perceptions and intentions: A study of agency, communality and trust. In *Congrès de l'Association Française du Marketing.*

Islind, A. S., Óskarsdóttir, M., Smith, S. Á., & Arnardóttir, E. S. (2023). The friendly chatbot: Revealing why people use chatbots through a study of user experience of conversational agents. *14th Scandinavian Conference on Information Systems.* 6.

Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018, April). Convey: Exploring the use of a context view for chatbots. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1-6). https://doi.org/10.1145/3173574.3174042

Janssen, A., Cardona, D. R., Passlick, J., & Breitner, M. H. (2022). How to Make chatbots productive–A user-oriented implementation framework. *International Journal of Human- Computer Studies, 168,* 102921. https://doi.org/10.1016/j.ijhcs.2022.102921

Jiang, Y., Yang, X., & Zheng, T. (2023). Make chatbots more adaptive: Dual pathways linking human-like cues and tailored response to trust in interactions with chatbots. *Computers in Human Behavior,* 138, 107485. https://doi.org/10.1016/j.chb.2022.107485

Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39, 31–36. https://doi.org/10.1007/bf02291575

Kandpal, P., Jasnani, K., Raut, R., & Bhorge, S. (2020, July). Contextual chatbot for healthcare purposes (using deep learning). In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)* (pp. 625-634). IEEE.

Kim, Y., Seok, J., & Roh, T. (2023). The linkage between quality of information systems and the impact of trust-based privacy on behavioral outcomes in unmanned convenience store: Moderating effect of gender and experience. *Technological Forecasting and Social Change, 196,* 122852. https://doi.org/10.1016/j.techfore.2023.122852

Kiuchi, K., Otsu, K., & Hayashi, Y. (2023). Psychological insights into the research and practice of embodied conversational agents, chatbots and social assistive robots: a systematic meta-review. *Behaviour & Information Technology,* 1-41. https://doi.org/10.1080/0144929X.2023.2286528

Klein, K., & Martinez, L. F. (2022). The impact of anthropomorphism on customer satisfaction in chatbot commerce: an experimental study in the food sector. *Electronic commerce research,* 1-37. https://doi.org/10.1007/s10660-022-09562-8

Konya-Baumbach, E., Biller, M., & von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior,* 139, 107513. https://doi.org/10.1016/j.chb.2022.107513

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*(4), 730–735. https://doi.org/10.1037/1528-3542.7.4.730

Kuhail, M. A., Thomas, J., Alramlawi, S., Shah, S. J. H., & Thornquist, E. (2022, October). Interacting with a chatbot-based advising system: understanding the effect of chatbot personality and user gender on behavior. In *Informatics* (Vol. 9, No. 4, p. 81). MDPI.

Kumar, R., & Ali, M. M. (2020). A review on chatbot design and implementation techniques. *Int. J. Eng. Technol, 7*(11), 2791-2800.

Laakasuo, M., Palomäki, J., & Köbis, N. (2021). Moral uncanny valley: A robot's appearance moderates how its decisions are judged. *International Journal of Social Robotics, 13*(7), 1679-1688. https://doi.org/10.1007/s12369-020-00738-6

Lai, Y., Lioliou, E., & Panagiotopoulos, P. (2021). Understanding Users' switching Intention to AI-Powered Healthcare Chatbots. In ECIS.

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). Empathic robots for long-term interaction: evaluating social presence, engagement and perceived support in children. *International Journal of Social Robotics, 6,* 329-341. https://doi.org/10.1007/s12369-014-0227-1

Linder, C. (2020). *The Effects of a Healthcare Chatbots' Language and Persona on User Trust, Satisfaction, and Chatbot Effectiveness* (Doctoral dissertation, Clemson University).

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking, 21*(10), 625-636. https://doi.org/10.1089/cyber.2018.0110

Lowenthal, P. R. (2012). Social presence: What is it? How do we measure it? In *ProQuest LLC eBooks* (pp. 1–187). https://eric.ed.gov/?id=ED548698

Mara, M., Appel, M., & Gnambs, T. (2022). Human-like robots and the uncanny valley. *Zeitschrift für Psychologie,.230*(1), 33–46. https://doi.org/10.1027/2151- 2604/a000486

Mathur, M. B., & Reichling, D. B. (2009, March). An uncanny game of trust: social trustworthiness of robots inferred from subtle anthropomorphic facial cues. In *Proceedings of the 4th ACM/ IEEE international conference on Human robot interaction* (pp. 313-314).

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review, 20*(3), 709-734. https://doi.org/10.5465/amr.1995.9508080335

Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2022). An emerging theory of avatar marketing. *Journal of Marketing, 86*(1), 67-90. https://doi.org/10.1177/0022242921996646

Nadeem, W., Khani, A. H., Schultz, C. D., Adam, N. A., Attar, R. W., & Hajli, N. (2020). How social presence drives commitment and loyalty with online brand communities? the role of social commerce trust. *Journal of Retailing and Consumer Services, 55,* 102136. https://doi.org/10.1016/j.jretconser.2020.102136

Nißen, M., Rüegger, D., Stieger, M., Flückiger, C., Allemand, M., v Wangenheim, F., & Kowatsch, T. (2022). The effects of health care chatbot personas with different social roles on the client-chatbot bond and usage intentions: development of a design codebook and web- based study. *Journal of medical Internet research, 24*(4), e32630. https://doi.org/10.2196/32630

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service—findings from a questionnaire study. *Interacting with Computers, 31*(3), 317-335. ttps://doi.org/10.1093/iwc/iwz022

NV, G. R., Vanimireddy, R. T., Mothe, V. S. K., & Nenavath, A. N. (2023). Conversational AI Chatbot for HealthCare. In *E3S Web of Conferences* (Vol. 391, p. 01114). EDP Sciences.

Pawlik, V. P. (2021, November). Design matters! how visual gendered anthropomorphic design cues

moderate the determinants of the behavioral intention towards using chatbots. In *International-al Workshop on Chatbot Research and Design* (pp. 192-208). Cham: Springer International Publishing.

Paradeda, R. B., Hashemian, M., Rodrigues, R. A., & Paiva, A. (2016). How facial expressions and small talk may influence trust in a robot. In Social Robotics: *8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings 8* (pp. 169-178). Springer International Publishing.

Phiri, M., & Munoriyarwa, A. (2023). Health Chatbots in Africa: Scoping Review. *Journal of Medical Internet Research, 25,* e35573. https://doi.org/10.2196/35573

Pimentel, D., Kalyanaraman, S., Lee, Y. H., & Halan, S. (2021). Voices of the unsung: The role of social presence and interactivity in building empathy in 360 video. *New media & society, 23*(8), 2230-2254. https://doi.org/10.1177/1461444821993

Puvača, M., & Kunac, I. (2023, December). Chatbot in business. In *International Scientific Conference on Economy, Management and Information Technologies* (Vol. 1, No. 1, pp. 219-225).

Qureshi, O. A., Wilkins, S., & Iqbal, H. (2024, March). Chatbots: Can They Satisfy Customers in the Banking Sector?. In *BUiD Doctoral Research Conference 2023: Multidisciplinary Studies* (pp. 360-376). Cham: Springer Nature Switzerland.

Saglam, R. B., Nurse, J. R. C., & Hodges, D. (2021). Privacy concerns in Chatbot interactions: When to trust and when to worry. In *HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II 23* (pp. 391-399). Springer International Publishing.

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience, 11*(2), 88–95. https://doi.org/10.1080/21507740.2020.1740350

Sanjeed, V. K., Kim, M. G., & Wang, C. (2020). Examining the Effect of Chatbot Gender and Gender Congruence between a Chatbot and a Customer in a Banking Context. *Journal of Korea Service Management Society, 21*(5), 46–73. https://doi.org/10.15706/jksms.2020.21.5.003

Sanny, L., Susastra, A., Roberts, C., & Yusramdaleni, R. (2020). The analysis of customer satisfaction factors which influence chatbot acceptance in Indonesia. *Management Science Letters, 10*(6), 1225-1232. https://doi.org/10.5267/j.msl.2019.11.036

Schadelbauer, L., Schlögl, S., & Groth, A. (2023). Linking Personality and Trust in Intelligent Virtual Assistants. *Multimodal Technologies and Interaction, 7*(6), 54. https://doi.org/10.3390/mti7060054

Schaefer, K. E. (2016). Measuring trust in human robot interactions: Development of the "trust perception scale-HRI". In *Robust intelligence and trust in autonomous systems* (pp. 191- 218). Boston, MA: Springer US.

Schillaci, C. E., De Cosmo, L. M., Piper, L., Nicotra, M., & Guido, G. (2024). Anthropomorphic chatbots' for future healthcare services: Effects of personality, gender, and roles on source

credibility, user satisfaction, and intention to use. *Technological Forecasting and Social Change, 199,* 123025. https://doi.org/10.1016/j.techfore.2023.123025

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149,* 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Statista Research Department (2024, May 29). Use of chatbots on brand websites worldwide 2024, by age. Retrieved from https://www.statista.com/statistics/1468615/chatbots- age-brand-website-worldwide/

Sugumar, M., & Chandra, S. (2021). Do I desire chatbots to be like humans? Exploring factors for adoption of chatbots for financial services. *Journal of International Technology and Information Management, 30*(3), 38-77. https://doi.org/10.58729/1941-6679.1501

Tahvanainen, L., Tetri, B., & Ahonen, O. (2024, May). Exploring and Extending Human-Centered Design to Develop AI-Enabled Wellbeing Technology in Healthcare. In *Nordic Conference on Digital Health and Wireless Solutions* (pp. 288-306). Cham: Springer Nature Switzerland.

Tatai, G., Csordás, A., Kiss, Á., Szaló, A., & Laufer, L. (2003, September). Happy chatbot, happy user. In *International workshop on intelligent virtual agents* (pp. 5-12). Berlin, Heidelberg: Springer Berlin Heidelberg.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53 –55. https://doi.org/10.5116/ijme.4dfb.8dfd

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry:

current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry, 20*(3), 318–335. https://doi.org/10.1002/wps.20883

van der Goot, M. J., & Pilgrim, T. (2019, November). Exploring age differences in motivations for and acceptance of chatbot communication in a customer service context. In *International Workshop on Chatbot Research and Design* (pp. 173-186). Cham: Springer International Publishing.

Vanneste, B., & Puranam, P. (2021). Artificial intelligence, trust, and perceptions of agency. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3897704

Vlachogianni, P., & Tselios, N. (2023). Perceived Usability Evaluation of Educational Technology Using the Post-Study System Usability Questionnaire (PSSUQ): A Systematic Review. *Sustainability, 15*(17), 12954. https://doi.org/10.3390/su151712954

Wang, X., Lin, X., & Shao, B. (2023). Artificial intelligence changes the way we work: A close look at innovating with chatbots. *Journal of the Association for Information Science and Technology, 74*(3), 339-353. https://doi.org/10.1002/asi.24621

Zhang, J., Lu, J., Wang, X., Liu, L., & Feng, Y. (2024). Emotional expressions of care and concern by customer service chatbots: Improved customer attitudes despite perceived inauthenticity. *Decision Support Systems, 186,* 114314. https://doi.org/10.1016/j.dss.2024.114314

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020). Towards Developing Trust-Supporting Design Features for AI-Based Chatbots in Customer Service. In ICIS (Vol. 2020, pp. 1-9).

Zogaj, A., Mähner, P. M., Yang, L., & Tscheulin, D. K. (2023). It'sa Match! The effects of chatbot anthropomorphization and chatbot gender on consumer behavior. *Journal of Business Research, 155,* 113412. https://doi.org/10.1016/j.jbusres.2022.113412

Zouzou, M. C., Benkhelifa, E., Kholidy, H., & Dyke, D. W. (2023, June). Multi-Context-aware Trust Management framework in Social Internet of Things (MCTM-SIoT). In *2023 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)* (pp. 99-104). IEEE.

Appendix A

Pre-test slides

Appendix B

Summary of pre-tests

Pre-test 1

Gender: Female

Age: 23

Nationality: Dutch

Can you imagine these bots as a financial/medical advisor? Yes

Favourites:
- Woman smiling with blazer (glasses don't matter)
- Woman neutral with blazer, then glasses are better
- Smiling and neutral medical are both preferred with stethoscope (glasses are nice, but less important than stethoscope)
- Smiling financial man is preferred with tie and glasses, neutral financial man is also preferred with tie but without glasses
- Medical man is preferred with stethoscope, neutral with glasses plus tie

Videos:

- Good that chatbot reassures about confidentiality
- "fine" is not a good word → a bit harsh
- Video speed is alright, good that bot takes time to answer

Feedback/remarks:

- Conversation is not completely realistic → a doctor may ask more questions
- Provide more information, e.g., "here is a pdf"
- Financial bot asks for information too soon, should ask for more information before planning

Pre-test 2

Gender: Female

Age: 24

Nationality: Dutch

Can you imagine these bots as a financial/medical advisor? Yes

Favourite:

- Slide 1: blazer + glasses
- Slide 2: blazer with and without glasses
- Slide 3: stethoscope without glasses
- Slide 4: stethoscope without glasses
- Slide 5: tie with glasses
- Slide 6: tie without glasses
- Slide 7: with stethoscope and glasses
- Slide 8: with stethoscope but without glasses

Videos:

- Less realistic; empathy is lacking
- Delay is good
- Language is very formal and professional
- Nice: introduction by name
- Finance does not ask about goal or further questions
- User may be less active (tick boxes instead of writing answers)
- Too specific terms in financial context
- Chatbots appear to be empathic
- One reply from Luke is too long to read before user answers

Feedback/remarks:

- Add link to confidentiality statement
- Less professional without glasses
- Blazer is more professional
- Stethoscope good for medical context
- Tie is more professional
- Glasses matter less in serious condition
- For women, hair is influential
- Stethoscope without tie is enough

Pre-test 3

Gender: Female

Age: 23

Nationality: German

Can you imagine these bots as a financial/medical advisor? Yes, for medicine

Favourite:

- Slide 1: with glasses and blazer
- Slide 2: with glasses and blazer
- Slide 3: with stethoscope and glasses
- Slide 4: with stethoscope and glasses
- Slide 5: with glasses and tie
- Slide 6: with glasses
- Slide 7: with stethoscope and glasses (tie is nice but less important)
- Slide 8: with stethoscope and glasses

Videos:

- Reassuring
- Professional
- Video speed very fast for long answers
- Smile has an effect
- Financial guy is more professional
- Medical bot is more appealing

Feedback/remarks:

- Glasses are more important in financial context
- Financial guy looks more like an expert

Pre-test 4

Gender: Male

Age: 24

Nationality: German

Can you imagine these bots as a financial/medical advisor? Only partly

Favourite:

- Slide 1: with blazer and glasses
- Slide 2: with blazer and glasses
- Slide 3: stethoscope and glasses
- Slide 4: stethoscope without glasses
- Slide 5: tie and glasses
- Slide 6: tie (glasses less important)
- Slide 7: stethoscope with glasses
- Slide 8: stethoscope without glasses, tie is not good for medical

Videos:

- Financial conversation is less professional → too abrupt
- Video speed of medical video is adequate

Feedback/remarks: -

Pre-test 5

Gender: Male

Age: 29

Nationality: German

Can you imagine these bots as a financial/medical advisor? slightly

Favourite:

- Blazer and glasses are always a nice combination
- Glasses are less important in medical condition
- Stethoscope is very good, tie is nice but stethoscope is enough

Videos:

- Finance video too fast; some sentences too long
- Medical video very relatable

Feedback/remarks:

- Reassuring the user is great → makes a good impression and is relatable

Pre-test 6

Gender: Female

Age: 60

Nationality: German

Can you imagine these bots as a financial/medical advisor? The financial one

Favourite:

- Slide 1: blazer without glasses
- Slide 2: shirt without glasses
- Slide 3: glasses without stethoscope
- Slide 4: without glasses, without stethoscope
- Slide 5: tie and glasses
- Slide 6: tie and glasses
- Slide 7: stethoscope and glasses
- Slide 8: stethoscope and glasses

Videos:

- Financial context should be less personal
- Avatars should introduce themselves
- Inform about confidentiality in the beginning
- Save "typing"

Feedback/remarks:

- Lose hair in woman is unprofessional
- Man is less realistic
- Man seems to be younger than woman
- All avatars better without stethoscope
- Avatars should not be this young (seem to be less experienced)
- Avatars could be more human-like

Appendix C

Questionnaire

Introduction

Welcome. Thank you for participating in this survey.
The study focuses on chatbots and it will take approximately 10 minutes to finish it.

This study will investigate user experience with chatbots and your personal opinion will be asked for. Participation is completely anonymous and there is no information that can be traced back to you. There are no harms or risks involved in participating in this study. The study is approved by the ethical committee of the University of Twente and participation is completely voluntarily.
If you have any questions, do not hesitate to contact the researcher l.v.simbach@student.utwente.nl or the ethics committee ethicscommittee-hss@utwente.nl.

Do you agree to take part in this study?

- Yes
- No

Please answer some general questions first.

Demographics

- Age
- Gender
- Nationality

Introductory questions

- Have you previously interacted with a chatbot?
- How often do you interact with a chatbot?
- How would you rate your experience with the interaction with chatbots in general?
- How comfortable are you in general with consulting a chatbot?

You will now watch a short video. Please pay close attention.

You can pause the video at any time. It may help to watch the video on full screen for better readibility.

Options 1 to 8

Please answer the following questions imagining you have been the user interacting with the chatbot.

Trust

- I would feel comfortable talking to the chatbot about a personal matter.
- I would feel in control over the interaction with the chatbot.
- I would feel that I as a user can trust the chatbot completely.
- I would feel that the chatbot is trustworthy.

Perceived ease of use

- It seems that it was simple to use this chatbot.
- It seems as if the chatbot was unnecessarily complex.
- Is seems that the chatbot was easy to use.
- It seems as if I would be able to interact with the chatbot without any efforts.

Empathy

- The chatbot understands the emotions of the user well.
- The chatbot was able to correctly respond to the users' emotions.
- The chatbot cared about the users' emotions.
- The chatbot was able to make the user feel comfortable.

Customer satisfaction

- I would be satisfied with the experience with the chatbot.
- I would be satisfied with the service the chatbot provided.
- I would recommend using this chatbot to my friends and family.
- I think the chatbot is adequately supporting the user.
- I think that the chatbot sufficiently solves the users' concern.
- I would have a positive feeling using the chatbot.

Competence

- The chatbot is skilled.
- The answers of the chatbot are helpful and thought-out.
- The chatbot responds considerately to the users' questions.
- The chatbots' performance was qualified.

Social presence

- I would feel emotionally connected to the chatbot.
- I would feel actively engaged in the conversation.
- I would feel a sense of connection with the chatbot.
- I would feel like I was talking to a real person.

Intention to use

- I think using the chatbot would be a good idea.
- I would use the chatbot if I have any concerns in the future.
- I think that I would like to use this chatbot frequently.
- I think that using the chatbot would be beneficial.

Thank you for answering the questions so far.

Manipulation check

- Was the chatbot male or female?
- Was the chatbot smiling or neutral?
- Was it a medical or financial conversation?

Thank you very much for participating in this survey!

Appendix D
AI statement

I have not made use of any AI tools for the finalisation of this thesis.