

**Chatbots as Social Actors:
Towards a Better Understanding of Social Responses to Text-
Based Conversational Service Agents**

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Stefanie Helene Klein
aus Leinfelden-Echterdingen

Tübingen
2024

1. Betreuerin:

Prof. Dr. Sonja Utz

2. Betreuer:

Prof. Dr. Dominik Papies

Tag der mündlichen Prüfung:

11.02.2025

Dekanin und Dekan:

Prof. Dr. Taiga Brahm und

Prof. Dr. Dominik Papies

1. Gutachterin:

Prof. Dr. Sonja Utz

2. Gutachter:

Prof. Dr. Stefan Mayer

Contents

List of Figures	vi
List of Tables.....	vii
1. General Introduction	1
1.1 Towards Machines as Communicators	3
1.2 Social Responses to Media Agents	4
1.3 Research Aims and Research Questions	7
1.4 Dissertation Outline.....	11
2. Agent type, Responsive Verbal Cues, and Delayed Response Time.....	15
2.1 Introduction	16
2.2 Related Research and Theoretical Background	16
2.2.1 User Responses to Chat Agents.....	16
2.2.2 Agent Type: Chatbot vs. Human.....	18
2.2.3 Responsiveness.....	19
2.2.4 The Interplay Between Agent Type and Responsiveness	19
2.2.5 Underlying Processes of Agent Type and Responsiveness Effects	20
2.2.6 Overview of the Current Studies	22
2.3 Study 1.....	22
2.3.1 Method.....	22
2.3.2 Results	23
2.3.3 Discussion	24
2.4 Study 2.....	24
2.4.1 Method.....	24
2.4.2 Results	24
2.4.3 Discussion	25
2.5 Study 3.....	25
2.5.1 Method.....	25
2.5.2 Results	25
2.5.3 Discussion	26
2.6 Additional Analyses: Pooled Data	26
2.7 General Discussion.....	30
2.7.1 Limitations and Future Research.....	33
2.8 Conclusion.....	33

2.9	Additional Results: The Impact of Response Time (Study 3)	35
3.	Free Text Interaction and Social Error Responses	39
3.1	Introduction	40
3.2	Theoretical Framework	42
3.2.1	Social Responses to Chatbots	42
3.2.2	Interaction Mechanism and User Responses to Chatbots	44
3.2.3	Responses to Errors and User Acceptance	45
3.2.4	The Interplay of Interaction Mechanism and Error Response	46
3.2.5	Underlying Processes	46
3.3	Method	48
3.3.1	Overview	48
3.3.2	Participants	48
3.3.3	Procedure	48
3.3.4	Stimuli and Manipulation of Independent Variables	49
3.3.5	Measures	52
3.4	Results	53
3.4.1	Manipulation Checks	53
3.4.2	Hypothesis Testing	54
3.4.3	Additional Analyses	60
3.5	Discussion	61
3.5.1	Theoretical Implications	62
3.5.2	Practical Implications and Outlook	63
3.5.3	Limitations and Future Research	65
3.6	Conclusion	65
4.	A Meta-Analysis of the Effects of Social Cues on Social Responses	67
4.1	Introduction	68
4.2	Main Effects of Human-Likeness	69
4.3	Moderators of Human-Likeness Effects	71
4.3.1	Sample Characteristics	71
4.3.2	Agent Characteristics	72
4.3.3	Context Characteristics	73
4.3.4	Method Characteristics	73
4.4	Method	74
4.4.1	Literature Search and Selection	74

4.4.2	Effect Size Calculation	76
4.5	Results	78
4.5.1	Main Effects	78
4.5.2	Publication Bias.....	80
4.5.3	Moderator Effects.....	81
4.6	Discussion	90
4.6.1	Implications for Theory and Research	92
4.6.2	Practical Implications	93
4.6.3	Limitations and Future Research.....	94
4.7	Conclusion.....	95
5.	General Discussion	97
5.1	Main Conclusions and Theoretical Implications.....	98
5.2	Strengths.....	103
5.3	Methodological Limitations	105
5.4	Practical Implications	106
5.5	Outlook and Future Research Questions.....	108
5.6	Conclusion.....	110
	References	111
	Appendix	139
A	Appendix Chapter 2.....	140
B	Appendix Chapter 3.....	155
C	Appendix Chapter 4.....	166
	Summary.....	197
	Zusammenfassung.....	198

List of Figures

Figure 1	Overview of the Dissertation.....	12
Figure 2	Parallel Multiple Mediator Model for the Effect of Responsiveness on Likability (Pooled Data).....	29
Figure 3	Parallel Multiple Mediator Model for the Effect of Responsiveness on Warmth (Pooled Data).....	29
Figure 4	Parallel Multiple Mediator Model for the Effect of Responsiveness on Satisfaction (Pooled Data).....	29
Figure 5	Research Framework.....	47
Figure 6	Translated Exemplary Screenshots of Button (left) and Free Text Chatbots.....	51
Figure 7	Interaction Plots of Error Response and Interaction Mechanism on Study Variables.....	57
Figure 8	Mediation Results of Interaction Mechanism on Satisfaction, Intention to Use, and Emotional Connection with the Company via Perceived Warmth.....	58
Figure 9	Mediation Results of Interaction Mechanism on Satisfaction, Intention to Use, and Emotional Connection with the Company via Perceived Competence.....	59
Figure 10	Analytical Framework.....	74
Figure 11	Summary Forest Plot.....	79
Figure 12	Contour-Enhanced Funnel Plots.....	80
Figure 13	Influence of Sample-Related Moderators on Human-Likeness Effects.....	82
Figure 14	Influence of Agent-Related Moderators on Human-Likeness Effects.....	83
Figure 15	Influence of Context-Related Moderators on Human-Likeness Effects.....	84
Figure 16	Influence of Method-Related Moderators on Human-Likeness Effects.....	85

List of Tables

Table 1	Two-Way ANCOVA Statistics and Effect Sizes for Study Variables (Pooled Data).....	27
Table 2	Adjusted Means and Effect Sizes for Study Variables (Pooled Data)	28
Table 3	Means, Standard Deviations, and Test Statistics in Study Variables by Response Time (Study 3)	36
Table 4	Three-way ANOVA Statistics and Effect Sizes for Study Variables (Study 3) ...	37
Table 5	Descriptive Statistics, Correlations, and Alpha Values for Study Variables	53
Table 6	Results of Two-Way Analyses of Variance in Study Variables	55
Table 7	Means and Standard Deviations of Interaction Mechanism and Error Response Main Effects on Study Variables.....	56
Table 8	Overview of Results	60
Table 9	Overview of Moderator Variables.....	77
Table 10	Meta-Analytic Results and Heterogeneity Assessment.....	79
Table 11	Subgroup Analyses for Overall Social Responses ($k = 800$).....	86
Table 12	Meta-Regression Results	89

1. General Introduction

In our everyday interactions with organizations, conversational agents (CA) based on artificial intelligence¹ (AI) provide us with services like answering our questions, recommending and selling us products, handling our complaints, or even lending us emotional support (Jacobson & Gorea, 2023; W. Liu & Yao, 2023). Driven by substantial advances in AI-based technologies like natural language processing (NLP) and, more recently, large language models (LLM), as well as the widespread availability of the internet and mobile devices like smartphones, access to and use of AI-based CAs is on the rise (Dale, 2016; OpenAI, 2023; Poushter et al., 2024; Wong & Bottorff, 2023). An illustrative example is the LLM-powered chatbot ChatGPT, which was launched in November 2022 and, according to OpenAI, is now used by 200 million people per week (Reuters, 2024). The current hype around AI-based CAs naturally focuses on their potential to facilitate people's lives or perform tasks more efficiently or more accurately than humans (Stein et al., 2023). However, the public has also expressed concerns about data privacy, lack of transparency, and the reproduction of biases (Abercrombie et al., 2023; Ischen, Araujo, Voorveld, et al., 2020). Given the disruptive nature of the underlying technologies, the inherent risks, and the fact that AI-based CAs have found their way into almost every corner of our daily lives, understanding their acceptance and adoption among end users is crucial.

Particularly CAs that interact with users via text (“chatbots”)² are increasingly used by organizations to enhance or replace human agents in customer service (Liu-Thompkins et al., 2022), that is, “the assistance and support provided by a brand to its customers before, during, and after they make a purchase or use a product or service” (Ferraro et al., 2024, p. 550). Chatbots are employed in various contexts, including e-commerce, finance, insurance, education, and tourism, with the global chatbot market predicted to expand from \$5 billion in 2023 to \$27 billion in U.S. dollars by 2030 (Grand View Research, 2023). Users can receive real-time support and conveniently access chatbots via various platforms and devices (Følstad et al., 2018). AI-driven chatbots create new opportunities for service providers to build and

¹ Throughout this dissertation, AI is considered as an umbrella term for technologies that aim to simulate human intelligence in machines, e.g., text-based conversational agents (Gunkel, 2020; McCarthy, 2007; Stein et al., 2023; Stewart et al., 2020).

² Chatbots can be defined as CAs that communicate with users via natural language (Allouch et al., 2021; Shawar & Atwell, 2007). While the first chatbots were mostly text-based, advances in speech technology and the subsequent development of voice assistants such as Siri and Alexa have led to the conceptual distinction between text-based CAs, now commonly termed chatbots, and voice-based CAs (Allouch et al., 2021). In this dissertation, the term chatbot refers to text-based CAs.

sustain relationships with consumers (M.-H. Huang & Rust, 2021b; Steinhoff et al., 2019), and they have also been shown to contribute positively to firm value (Fotheringham & Wiles, 2022).

Both users and service providers share a vested interest in successful service encounters, i.e., that the user's problem is solved to their satisfaction in good time (Dixon et al., 2014). Yet, the success of service encounters with chatbots is frequently jeopardized by factors related to insufficient design or technology, resulting in conversational breakdowns and user frustration (Benner et al., 2021; Brendel et al., 2023). Unsurprisingly, less than 10% of consumers recently surveyed by Gartner said they had consulted a chatbot the last time they needed customer support, and only a quarter of those people would contact that chatbot again (LoDolce & Brackenbury, 2023). Even when interacting with apparently proficient LLM-based chatbots, users often experience confusion and frustration because the chatbot fails to understand the user's message or gives incorrect answers, so-called *hallucinations* (Khennouche et al., 2024; Stokel-Walker & Van Noorden, 2023). Chatbots that do not live up to the expectation of a seamless service encounter can have detrimental consequences for users and service providers, for instance, in terms of social responses related to trust (Seeger & Heinzl, 2021), word-of-mouth (Yu et al., 2024), purchase intention, and firm evaluation (Crollic et al., 2022). So, do users prefer to interact with a human agent in service encounters, after all?

This dissertation aims to improve our understanding of peoples' social responses to chatbots. However, it is still unclear if and when users favor interacting with a chatbot or a human agent in a service encounter (G. Huang & Wang, 2023; Van der Goot, 2022). Thus, before comparing different chatbot versions, this dissertation first aims to investigate whether users' social responses to chatbots and human chat agents differ by building on the machine heuristic of the Modality-Agency-Interactivity-Navigability (MAIN) model (Sundar, 2008; Sundar et al., 2019). In the next step, this dissertation focuses on chatbots in service encounters. Drawing on the Media Are Social Actors (MASA) paradigm (Lombard & Xu, 2021), I investigate how chatbots need to communicate to improve people's service encounter experience. Previous research suggests that humanizing chatbots through the use of social cues can mitigate negative chatbot experiences (e.g., Brendel et al., 2023; De Sá Siqueira et al., 2023), and enhance user engagement and the relationship between service providers and service users (e.g., Adam et al., 2021; Araujo, 2018; Bergner et al., 2023; Nguyen et al., 2023). Social cues related to many design aspects of the chatbot, including verbal, visual, invisible, and interaction cues (Feine et al., 2019; Zierau et al., 2020), can potentially influence the responses of users by leading them to apply scripts from interpersonal interaction to their interactions with chatbots (Lombard & Xu, 2021). This thesis concentrates on four potentially powerful aspects

of human-likeness that have not been researched sufficiently: responsiveness (Reis & Clark, 2013), delayed responses (Kaptein et al., 2011), social error responses (P. H. Kim et al., 2004), and free text interaction (Jain et al., 2018). Although human-likeness is generally seen as beneficial, human-like chatbots can have negative or no effects on social responses under certain circumstances (e.g., Ciechanowski et al., 2019; Crollic et al., 2022; Gnewuch et al., 2022). Moreover, not all types of social cues can be investigated in one dissertation. Therefore, this dissertation includes a meta-analysis to gain generalizing insights into the relationship between social cues and users' social responses towards text-based chatbots and potential moderating factors. Before explaining the theoretical framework of this dissertation, the next section provides relevant background information on the research field in which this dissertation is situated.

1.1 Towards Machines as Communicators

Communication research has traditionally focused on human communication, with communication theory addressing “the media, modalities, and messages by which humans exchange, reflect on, and enact different perspectives of reality” (Jensen, 2008, p. 1). Until recently, computer-mediated communication (CMC), i.e., communication between humans facilitated by computers, held a central position in communication research (Gunkel, 2020; Kerr & Hiltz, 1982). However, the rise of AI-based technologies like NLP and neural networks has led to the development of what Gambino et al. (2020) coined “media agents” (p. 73). A media agent is defined as “any technological artifact that demonstrates sufficient social cues to indicate the potential to be a source of social interaction” (p. 73). Chatbots and voice assistants, avatars in virtual environments, smart devices like smart watches and data glasses, and embodied social robots that communicate with users are examples of media agents (Fong et al., 2003; Gambino et al., 2020).

Chatbots have existed since the 1960s, when ELIZA, the first text-based dialogue system, was invented by Joseph Weizenbaum. ELIZA was programmed to be a psychotherapist and responded to a user with a question based on the keywords in the user's query (Weizenbaum, 1966). At the time, the goal was to make human-chatbot conversations as undistinguishable from human conversations as possible (Turing, 1950). For instance, the Loebner Prize was created to recognize the most human-like chatbots (Mauldin, 1994). Initially designed to simulate human conversation, chatbots have evolved considerably. Milestones include the development of the artificial intelligence markup language (AIML)-based chatbots ALICE and Kuki regularly winning the Loebner Prize in the 2000s and 2010s (R. S. Wallace, 2009), and the subsequent development of voice-based conversational agents like Apple's Siri,

IBM's Watson, Amazon's Alexa, and Google Assistant, culminating in more recent innovations such as the LLM-based chatbots Replika and ChatGPT (Grudin & Jacques, 2019; OpenAI, 2024a; Zemčik, 2019).

In the late 2010s, a new communication paradigm called Human-Machine Communication (HMC) emerged from related subfields and disciplines, including Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) (Fortunati & Edwards, 2020, 2021; Guzman, 2018). Andrea Guzman, an HMC pioneer, defines HMC as “the creation of meaning between humans and machines, with technology theorized as a communicator, a subject with which people communicate, instead of a channel through which humans interact with one another” (p. 2). This definition emphasizes the departure from traditional CMC, where humans communicate with other humans *through* media, towards interactions between humans and media agents, i.e., humans communicating *with* media agents (Fortunati & Edwards, 2020; Guzman, 2018). Media agents are positioned as “communicative subject[s] directly exchanging messages with people in a meaning-making process” (Guzman & Lewis, 2020, p. 74). This dissertation takes on an HMC approach and thus views communication not only as a process between humans, and technology not merely as a tool. Instead, this dissertation considers text-based chatbots as communicators that can engage in meaningful social interactions with humans (Guzman, 2018; Guzman & Lewis, 2020).

1.2 Social Responses to Media Agents

Two theoretical approaches have heavily influenced research on communication between media agents and humans: the Computers Are Social Actors (CASA) paradigm (Nass et al., 1994; Nass & Moon, 2000) and the Modality-Agency-Interactivity-Navigability (MAIN) model (Sundar, 2008). Both approaches assume that mainly heuristics (i.e., simple, low-effort rules by which people make decisions) determine how people respond to media agents (Dehnert & Mongeau, 2022).

The CASA paradigm originally stems from the Media Equation, which states that individuals tend to mindlessly attribute human characteristics to media technologies, interacting with them as if they were genuine social actors (Reeves & Nass, 1996). In their seminal book, Reeves and Nass (1996) experimentally showed that humans apply scripts (i.e., mental representations of fixed procedures for everyday situations (Schank & Abelson, 1975)) from interpersonal communication to their interactions with media technologies. They concluded that media technologies (i.e., desktop computers and television sets) are not just tools but „full participants in our social and natural world“ (p. 251). Compared to the Media Equation that focused on various sorts of then-new media technologies (e.g., desktop computers), the CASA

paradigm, with its ideas sometimes referred to as Social Response Theory (Diederich et al., 2019), focuses more specifically on interactions with digital technologies that have the potential to communicate with users in an agentic way (Gambino et al., 2020; Nass & Moon, 2000). Whether people react socially to agents depends on three conditions present in almost every computer (application) (Nass & Moon, 2000; Trepte et al., 2021): 1) language, i.e., whether the computer uses words to communicate, 2) interactivity, i.e., whether the computer's responses depend on a user's previous input, and 3) whether the computer takes on a role normally performed by a human. Despite its popularity, CASA has recently come under review. For instance, by replicating Nass et al. (1994)'s study on politeness toward computers, Heyselaar (2023) showed “that participants no longer interact with desktop computers as if they are human” (p. 1). In addition, since “people have changed, technologies have changed, and the way people interact with technologies has changed” (Gambino et al., 2020, p. 1), calls for adapting or extending the CASA paradigm emerged (Gambino et al., 2020; E.-J. Lee, 2024; Lombard & Xu, 2021; Van der Goot & Etzrodt, 2023).

One current extension of CASA suitable for investigating interactions between humans and media agents like text-based chatbots is the MASA paradigm (Lombard & Xu, 2021; Xu, Liu, et al., 2023). Compared to CASA, MASA emphasizes the impact of social cues—“features salient to observers because of their potential to convey useful information” (Fiore et al., 2013, p. 2)—on users' responses to media agents. Previous research has categorized social cues with regard to the perceptible design elements that represent them (Feine et al., 2019; Zierau et al., 2020). Feine et al. (2019)'s taxonomy includes three categories of social cues that are applicable to text-based chatbots³: Verbal cues refer to written or spoken expressions like greetings. Visual cues encompass cues that users can see, e.g., a human-like picture, and invisible cues are cues that users cannot see, e.g., dynamic response times (Feine et al., 2019). This taxonomy can be extended by adding cues that relate to the interaction or, in other words, social dynamics (Fogg, 2003), e.g., the chatbot's interaction mechanism (Zierau et al., 2020).

Social cues can convey social signals that refer to how the observer interprets the social cues of the sender media agent (Fiore et al., 2013) or perceives it as a social actor (Lombard & Xu, 2021), e.g., personality or social presence. Finally, social signals may elicit social responses

³ A fourth category, auditory cues, includes audible voice features (Feine et al., 2019), and therefore does not apply to text-based chatbots, the subject of this dissertation.

from users toward media agents⁴. Responses are social if they are “similar to reactions shown during interactions with other human beings” (Krämer, 2005, p. 443). These responses can be cognitive (e.g., perceived intelligence), affective (e.g., perceived joy), or behavioral (e.g., continuance intention, information disclosure). While CASA assumes that even minimal social cues, i.e., language use, interactivity, and taking on a *human* role, can evoke social responses (Nass et al., 1994; Nass & Moon, 2000), MASA introduces a hierarchy of social cues (Lombard & Xu, 2021). Primary social cues, such as eye gaze and human-like movement, “are most salient and central to humans’ perception of socialness” (p. 32) and sufficient to trigger social responses. Secondary social cues, like the human-like use of language, are “less salient and less central to humans’ perception of socialness [and] neither sufficient nor necessary to evoke social responses” (p. 32f.). An explanation for how social cues impact social responses is anthropomorphism (Xu, Liu, et al., 2023), defined as “the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions” (Epley et al., 2007, p. 864). Researchers have long assumed that anthropomorphism is mindful; people hold the “thoughtful, sincere belief that the object has human characteristics” (Nass & Moon, 2000, p. 93). However, Y. Kim & Sundar (2012) experimentally confirmed that social cues can also elicit mindless evaluations of a human-like character on a website, i.e., in terms of perceived friendliness, sociability, likability, and personalness (Y. Kim & Sundar, 2012).

The MASA paradigm will be the overarching theoretical framework for this dissertation. As a complement, the MAIN model, particularly the *machine heuristic*, will be used to explain potential differences in how people perceive media agents and humans (E.-J. Lee, 2024; Sundar, 2008; Sundar & Kim, 2019). The MAIN model states that perceived technological affordances, as cues in the technological interface, trigger heuristics that influence users’ interaction experiences and their attitudinal and behavioral responses toward the interface (Sundar, 2008, 2020). Four affordances (i.e., perceived opportunities for action offered by an interface (Norman, 1999)) are at the model's center: modality, agency, interactivity, and navigability⁵ (Sundar, 2008). This dissertation focuses on the agency

⁴ Social responses are sometimes referred to as persuasive outcomes (Dehnert & Mongeau, 2022; G. Huang & Wang, 2023), emphasizing the goal of a communicator, e.g., a media agent, to form or change the receivers’ attitudes, perceptions, and behaviors (Miller, 2013). However, in this dissertation, the term social responses will be used as it is a crucial concept of the MASA paradigm.

⁵ Modality refers to how information is communicated, e.g., through voice or text. Interactivity focuses on what choices the interface provides the user with and whether the interface is responsive. Navigability entails cues that enable the user to navigate the interface smoothly (Sundar, 2008).

affordance, which refers to who users perceive as the communicator of information (Sundar, 2008). Who or what users perceive as the source of information and how they evaluate this source depends on the cues of the interface (e.g., machine-like appearance and communication). The *machine heuristic*, “the rule of thumb that machines are more objective than humans” (Sundar & Kim, 2019, p. 2), conceptually refines the MASA paradigm by explaining people’s different responses towards different sources of information, for example, machines like text-based chatbots and human agents (E.-J. Lee, 2024). In the next section, building on the theoretical framework and previous empirical work, I present the research aims of this dissertation and the research questions it attempts to answer.

1.3 Research Aims and Research Questions

This dissertation sets out to investigate the differences in users’ social responses to text-based service chatbots compared to human service agents and other versions of text-based chatbots. First, this dissertation aims to examine how the alleged identity of an agent, i.e., the agent type (chatbot vs. human agent), influences users’ social responses to an agent. Second, by moving on to comparing different chatbot versions, this dissertation seeks to investigate to what extent social cues influence users’ social responses, individually and from a meta-perspective. Third, this dissertation aims to shed light on the underlying mechanisms of the effects of social cues on users’ responses by considering potential mediators, i.e., social signals, like users’ perceptions of the agent and their interaction experience. The fourth goal of this dissertation is to determine to what extent and under what circumstances (i.e., varying characteristics of sample, agent, context, and method) social cues improve users’ social responses on a meta-level.

Understanding users’ social responses to chatbots and human agents

Whether people prefer interacting with a human or a chatbot might depend on several factors. One explanation is provided by the *machine heuristic* as part of the agency affordance in the MAIN model (Sundar, 2008; Sundar & Kim, 2019). This heuristic is triggered when people perceive a media agent as machine-like due to the cues present on the interface (Sundar & Kim, 2019). Consequently, people perceive this technology as objective, resulting in potentially improved social responses, i.e., more favorable judgments of credibility, attitudes, or behaviors toward the machine and the service (Sundar et al., 2019). For instance, prior research has shown that people were more open to telling their payment information to a machine agent than a human agent when buying plane tickets (Sundar & Kim, 2019). However, machine-like cues can also trigger perceptions of cold, inflexibility, and lack of emotion, particularly when a machine performs a task requiring an agent to adequately respond to the

needs of the communication partner. Prior research on the impact of agent type on social responses primarily looked at task-based customer service scenarios (e.g., Go & Sundar, 2019; Lou et al., 2022). In this dissertation, whether and how people's social responses towards a chatbot or a human agent differ is discussed in the context of developmental study advising, an inherently human task (Mottarella et al., 2004; Sundar, 2020). Thus, a reverse or negative machine heuristic could be triggered leading people to prefer a human agent (Sundar, 2020).

A comprehensive set of social responses potentially influenced by agent type is investigated, including a) cognitive, b) affective, and c) behavioral social responses and overall evaluations of the agent. This distinction stems from previous research on attitudes, particularly the idea that people evaluate an object "based on cognitive, affective, and behavioral information" (Maio et al., 2019, p. 4), i.e., what people think and how they feel about an object, and what they (intend to) do. Perceptions of the agent (i.e., likability, warmth, intelligence, and competence) constitute cognitive social responses. The affective social response is the enjoyment perceived during the interaction. The behavioral component is represented by the intention to use the agent in the future. Overall evaluations include users' general attitude towards the agent and satisfaction with the agent's performance. The first research question is:

RQ1. To what extent does the identity of an agent taking over a human task influence cognitive, affective, and behavioral social responses?

Understanding the impact of human-likeness in the form of verbal, invisible, and interaction cues on social responses to chatbots

With the second aim of this dissertation, the focus is reoriented to comparing chatbots that differ with respect to their human-likeness, i.e., the social cues they display. This dissertation experimentally investigates the impact of four social cues, i.e., two verbal cues, one invisible cue, and one interaction cue, on users' cognitive, affective, and behavioral social responses as well as their overall evaluations of the interaction. The first verbal cue this dissertation addresses refers to language that expresses responsiveness, referring to partners' mutual attention and support conveyed through relevant and adequate responses (Davis & Perrowitz, 1979; Reis & Clark, 2013). Responsiveness in relationships between humans has been shown to correlate positively with well-being (Reis & Clark, 2013) and in service contexts with customer satisfaction and trust (De Ruyter & Wetzels, 2000).

The second verbal cue this dissertation considers refers to the chatbot's response to errors it has made. Social error responses include the chatbot apologizing for not understanding the user's request, i.e., acknowledging responsibility, expressing regret, and admitting that it is still learning (P. H. Kim et al., 2004). Apologies are crucial in human-to-human interactions,

for instance, for future relationship expectations among friends (Hodgins & Liebeskind, 2003) and restoring trust in the workplace (Walfisch et al., 2013).

The invisible social cue this dissertation considers refers to the chatbot's response time, i.e., how much time passes until a user receives a response from an agent (Moon, 1999). Users generally prefer an immediate answer from a technological system (Hoxmeier & DiCesare, 2000). However, dynamically delayed response times, as typically encountered in computer-mediated interactions between humans may be beneficial as they can increase feelings of social presence, which may mediate response time effects on social responses (Gnewuch et al., 2018).

The interaction cue of interest refers to whether users interact with the chatbot primarily via free text or predefined buttons. Free text interaction might increase social responses as computer-mediated communication between humans often also occurs via free text chats (Jain et al., 2018). Responsive verbal cues, social error responses, dynamically delayed responses, and free text interaction are social cues that potentially make the chatbot interaction more human-like, leading users to apply scripts from interpersonal interaction to their interaction with the chatbot, in turn improving social responses (Lombard & Xu, 2021).

Besides the impact of individual social cues on social responses investigated in four experimental studies, the overall effect of social cues on a range of categories of social responses is also systematically analyzed from a meta-perspective. Research on the impact of social cues in chatbots has recently evolved rapidly, particularly since the starting point of this dissertation (Gopinath & Kasilingam, 2023). However, there is empirical evidence for ambiguous effects of human-likeness on social responses towards chatbots. For example, humanizing chatbots can increase frustration when interactions are flawed (Brendel et al., 2023), when customers are angry (Crolic et al., 2022), or when the chatbot avatar seems too familiar (Song & Shin, 2024). This dissertation thus aims to provide generalized insights into the overall effect on social responses as well as the individual effects on cognitive (i.e., perception, trust, rapport), affective (i.e., positive and negative affect), and behavioral outcomes through a meta-analysis. Prior meta-analyses on text-based chatbots have looked at predictors of intention to use service chatbots (Gopinath & Kasilingam, 2023), the effectiveness of chatbots on mental health (Abd-Alrazaq et al., 2020), and lifestyle outcomes (Singh et al., 2023). The impact of human-likeness has not been studied specifically for text-based chatbots but more so for more embodied agents like social robots (e.g., Blut et al., 2021; Roesler et al., 2021; Xu, Chen, et al., 2023). Yet, compared to more embodied agents that can communicate via other channels like speech, facial expressions, and gestures, text-based chatbots distinctly rely on a messaging-based interface (Araujo, 2018), which is why human-likeness might work

differently here. Therefore, one goal of this dissertation is to determine the overall effectiveness of social cues on social responses. The second research question is:

RQ2. To what extent does human-likeness in the form of social cues influence users' cognitive, affective, and behavioral social responses to chatbots?

Understanding the mechanisms that explain the influence of agent identity and human-likeness in the form of social cues on social responses

Social signals that entail how users perceive the agent or the interaction based on the agent's social cues may function as an intermediary link between agent type and social cues and social responses (Ischen, Araujo, van Noort, et al., 2020; Lombard & Xu, 2021; Sundar, 2020). This dissertation employs perceptions of the interaction experience, i.e., social presence, perceived dialogue, and feeling heard, as well as perceptions of the agent, i.e., perceived warmth and competence, as social signals that can potentially explain the relationship between social cues and users' social responses. Particularly social presence, "the sense of being with another" (Biocca et al., 2003, p. 468), is an established mediator in HMC research (Sundar, 2008; Van der Goot & Etzrodt, 2023). Perceived dialogue includes the notion of a dialogic back-and-forth conversation, which has been shown to increase an agent's perceived expertise and friendliness, as well as attitudes towards and intention to re-visit the website where the agent was placed (Go & Sundar, 2019; Sundar et al., 2016). Feeling heard is a relatively new concept that represents "the feeling that one's communication is received with attention, empathy, respect, and in a spirit of mutual understanding" (Roos et al., 2023, p. 5). Perceptions of the agent include perceived warmth and competence, which are central concepts that people apply in their evaluations of new actors (Fiske et al., 2002).

Human identity cues, responsive verbal cues, social error responses, and human-like free text interaction are social cues that potentially make the chatbot more human-like, improving perceptions of the agent and the interaction experience. Moreover, the above perceptions are likely positively associated with the investigated affective, cognitive, and behavioral social responses (Go & Sundar, 2019; Sundar, 2008, 2020). Users are likely to see chatbots as social actors and communicative subjects and thus potentially apply these concepts to them. Therefore, this dissertation aims to identify how verbal and interaction cues influence users' social responses. The third research question is:

RQ3. To what extent do perceptions of the interaction experience and the agent explain the effects of agent identity and social cues on cognitive, affective, and behavioral social responses?

Understanding the factors that influence the effects of social cues on users' social responses

The fourth aim relates to factors that moderate social cue effects on social responses. Prior research has shown mixed evidence regarding the impact of social cues on users' social responses—human-likeness might, in fact, influence user acceptance negatively in some cases (Ciechanowski et al., 2019; Crolic et al., 2022; Gnewuch et al., 2022). This dissertation sheds light on the factors that potentially determine the effectiveness of human-like social cues in text-based chatbots, including characteristics of the sample under investigation, the agent employed as the stimulus in the study, the context in which the study was placed, and methodological factors.

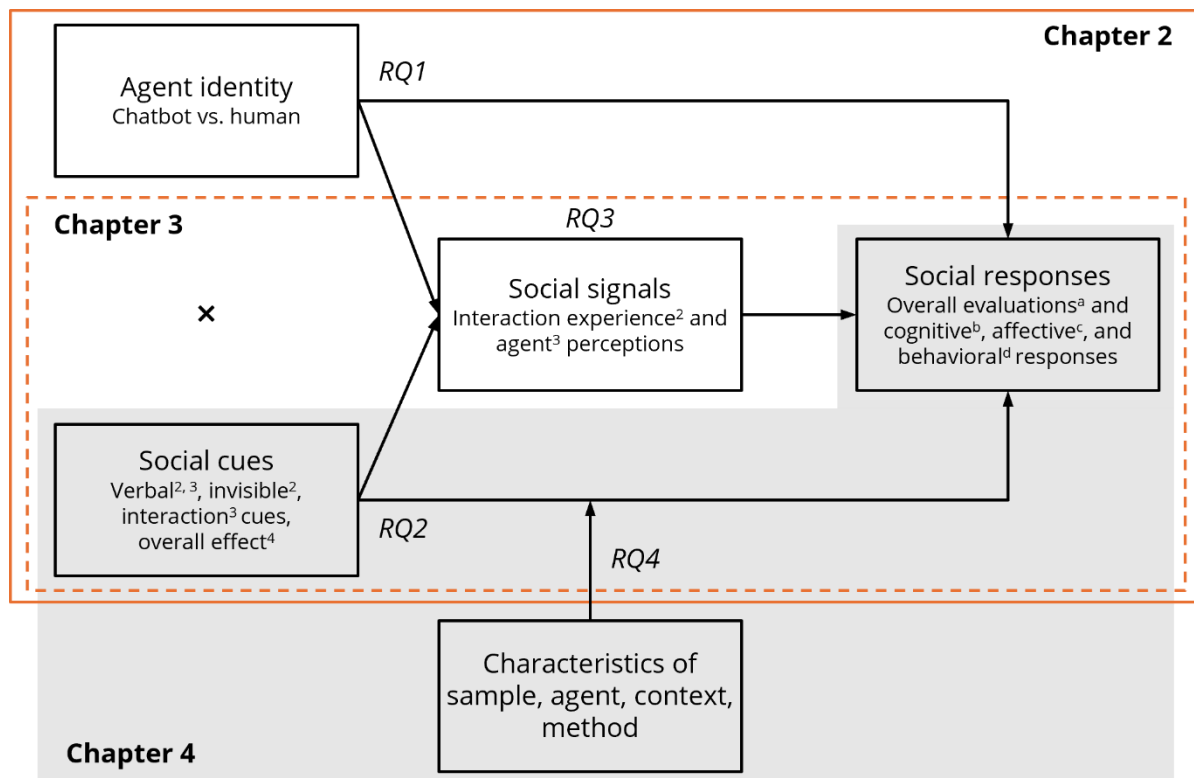
One set of potential moderators includes sample characteristics like gender (S. Kim et al., 2021; Pelau et al., 2021), age (Scopelliti et al., 2005; Straßmann et al., 2018), and if participants have used a chatbot before (Gnewuch et al., 2022). Next, specific features of the stimulus chatbot, such as its assigned gender, human-like representation, interaction mechanism, and the social cues used to implement human-likeness, might influence the effectiveness of human-likeness (Blut et al., 2021). Further, the interaction context and the nature of the chatbot's task might affect whether human-likeness can improve social responses (Lombard & Xu, 2021; Roesler et al., 2021). In addition, the effectiveness of human-likeness could also depend on method-related factors, such as publication status and whether participants watch prerecorded interactions or directly interact with a chatbot. The fourth research question is:

RQ4: Under what circumstances do social cues improve social responses to chatbots?

1.4 Dissertation Outline

Following the general introduction (Chapter 1), this dissertation comprises three empirical papers, presented in Chapters 2, 3, and 4. The papers presented in Chapter 2 and 3 are published, and the paper presented in Chapter 4 has been submitted to a peer-reviewed journal. Therefore, each chapter can be read individually. I provide an overview of the research questions this dissertation investigates in Figure 1.

Chapter 2, titled “Chatbot vs. Human: The Impact of Responsive Conversational Features on Users' Responses to Chat Advisors“, is a joint work with Sonja Utz. In contrast to prior work on task-based scenarios, it focuses on developmental study advising, a task that requires human skills. The chapter presents findings from three randomized online experiments wherein participants watched prerecorded chat interactions between a user and a study advisor.

Figure 1 Overview of the Dissertation

Note. ² in Chapter 2, ³ in Chapter 3, ⁴ in Chapter 4.

^a Chapter 2: attitude, satisfaction, Chapter 3: satisfaction, Chapter 4: attitude. ^b Chapter 2: likability, perceived intelligence, warmth, competence, Chapter 3: perceived warmth, competence, Chapter 4: perception, trust, rapport. ^c Chapter 2: perceived enjoyment, Chapter 3: emotional connection with the company, Chapter 4: positive and negative affect. ^d Chapters 2 and 3: intention to use, Chapter 4: behavioral outcomes.

It investigates RQ1–3 by examining the influence of agent identity (chatbot vs. human) and responsive verbal cues on affective, cognitive, and behavioral social responses and if perceptions of the interaction experience explain these effects, as illustrated by the orange box with a solid border in Figure 1. In addition, Chapter 2 explores the impact of invisible cues (i.e., dynamically delayed responses) on social responses.

Chapter 3, titled “How Interaction Mechanism and Error Responses Influence Users’ Responses to Customer Service Chatbots”, is a joint work with Dominik Papies and Sonja Utz. This chapter compared users’ responses to chatbots that exhibit different levels of verbal and interaction cues, as depicted by the dashed orange box in Figure 1. The chapter presents the results of a randomized online experiment wherein participants interacted with a customer service chatbot that makes mistakes. The chatbot was developed in collaboration with a company specializing in conversational AI. This chapter addresses RQ2 and RQ3 by investigating if free text interaction (i.e., interaction cue) and social error responses (i.e., verbal cue) improve agent perceptions of warmth and competence, user satisfaction, intention to use,

and emotional connection with the company. Additionally, it examines whether the social cues interact and whether perceived warmth and competence mediate the effects on satisfaction, intention to use, and emotional connection with the company.

Chapter 4, titled “The Effects of Human-Like Social Cues on Social Responses Towards Text-Based Conversational Agents—A Meta-Analysis”, is a single-authored project. The manuscript reports a meta-analysis of the effects of a chatbot’s social cues on users’ social responses, including their attitudes, perceptions, positive and negative affect, rapport, trust, and behavioral responses. While earlier meta-analyses have focused on human-robot interaction (Roesler et al., 2021) or interactions with AI-based service agents (Blut et al., 2021), this meta-analysis exclusively addresses text-based chatbots. Chapter 4 sets out to answer RQ2 and RQ4 by assessing the effect of social cues on social responses as well as by exploring sample-, agent-context- and method-related factors that may moderate this effect from a meta-perspective, as illustrated by the grey-shaded box in Figure 1. The meta-analysis follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). While Chapters 2 and 3 address the effects of selected individual social cues, Chapter 4 synthesizes the effects of text-based chatbots’ human-like social cues on social responses on a meta-level.

Chapter 5 contains a general discussion of the main findings as well as theoretical and practical implications, strengths and limitations, and ideas for future research.

2. Agent type, Responsive Verbal Cues, and Delayed Response Time

Statement of Contribution

Stefanie Helene Klein developed the research question, created the stimulus materials, conducted data collection, management, and analyses, and drafted the first version of the manuscript. Sonja Utz supervised the research project, gave feedback on conceptualization and methodology, and edited and revised the manuscript. This chapter presents a slightly modified version of the published article:

Klein, S. H., & Utz, S. (2024). Chatbot vs. human: The impact of responsive conversational features on users' responses to chat advisors. *Human-Machine Communication*, 8, 73–99. <https://doi.org/10.30658/hmc.8.4>

Additional preregistered analyses relating to the impact of response time on social responses that are not part of the published paper are presented in Chapter 2.9.

Abstract

As educational organizations increasingly consider supporting or replacing human chat advisors with chatbots, it is crucial to examine if users perceive a chatbot differently from a human. Chatbots' conversational features may signal responsiveness and thus improve user responses. To explore this, we conducted three online experiments ($N_{\text{total}} = 1,005$) using a study advising setting. We computed pooled data analyses because the individual study results did not provide clear support for our hypotheses. Results indicate that users prefer human agents regarding competence and intention to use but not perceived enjoyment. Responsiveness increased likability, warmth, and satisfaction. Perceptions of the interaction mediated the responsiveness effects. Our findings suggest that educational organizations can support their study advising departments with well-functioning chatbots without eliciting negative user responses.

Keywords: agent type, responsiveness, chatbot, user response, human-machine communication

2.1 Introduction

Chatbots are text-based automated agents interacting with users through natural language (Shawar & Atwell, 2007). They increasingly rely on artificial intelligence (AI-)based technologies, including large language models (LLMs) like GPT-3 (Kasneci et al., 2023). Already a part of many organizations' communication policies, AI-based chatbots are also increasingly used in educational contexts (e.g., for learning support or academic advising; Karrenbauer et al., 2021; Kasneci et al., 2023). Academic advising can be prescriptive, providing information about administrative issues, or developmental, aiming at defining and exploring study or career goals (Gordon, 1994; Mottarella et al., 2004). Developmental advising holds great potential for using AI-based chatbots but has received little attention in previous research (Meyer von Wolff et al., 2020). However, because developmental advising entails more than simply answering information requests, we deem it crucial to systematically investigate whether potential students prefer to be advised by a human or a chatbot (Sundar, 2020). Although developers have made significant technological progress in developing conversational agents, increasing chatbot acceptance among users remains a challenge (Neururer et al., 2018). Responsiveness, in the form of backchanneling cues, is a promising conversational feature that has been positively linked to organizational and relational outcomes in prior research on human-human (Davis & Perkowski, 1979; De Ruyter & Wetzels, 2000) and human-robot interactions (Birnbaum et al., 2016) but has hardly been studied in chatbots, as shown in a recent systematic review by van Pinxteren et al. (2020). We thus consider it important to examine the effects of chat agents' responsiveness on user responses. In three online vignette experiments, we aim to answer the research question: "To what extent do agent type (chatbot vs. human) and responsiveness influence users' responses to chat advisors?" Additionally, we look at the underlying mechanisms from three levels relating to different aspects of the interaction (i.e., the interaction in general, the dialogic nature, and the content of the conversation).

2.2 Related Research and Theoretical Background

2.2.1 User Responses to Chat Agents

Before turning to prior work on agent type and responsiveness, we want to introduce the user responses that form our dependent variables. Following previous work in human-agent interaction (e.g., Diers, 2020; S. Lee & Choi, 2017) and academic advising (e.g., Mottarella et al., 2004), we examine users' general attitude toward the way of communicating with an organization. To gain a deeper understanding, we look at specific cognitive components; for

example, the extent to which the advisor is perceived as likable (b), intelligent (c), warm (d), and competent (e). These perceptions are basic dimensions in evaluating new actors (Bartneck et al., 2009; Fiske, 2018). Likability and perceived intelligence are concepts stemming from human-robot interaction research. Warmth (i.e., good social intentions) and competence (i.e., the ability or expertise of the advisor) play a crucial role in evaluating human study advisors (Lou et al., 2022; Mottarella et al., 2004). Likability and warmth cover the social aspect of agent perception, while intelligence and competence are rather task-related (Bartneck et al., 2009; Fiske, 2018). We also capture users' affective and behavioral ratings of chat advisors to provide a comprehensive picture. Perceived enjoyment (f) constitutes the affective component. It refers to the extent to which interacting with a system is perceived as pleasurable and fun (Diers, 2020). We conceptualize satisfaction (g) as participants' perceived performance of the advisor they see in the vignette. In line with the Technology Acceptance Model, a prominent model to explain users' acceptance and usage of emerging technologies (Venkatesh, 2000), attitude, perceived enjoyment, and satisfaction are considered to be antecedents of user acceptance in terms of intention to use the way of communication with the organization (Diers, 2020; S. Lee & Choi, 2017). We also include the intention to use the communication medium (h), which, in turn, is considered to predict actual usage behavior (Venkatesh, 2000), but we were unable to measure this in our vignette studies.

We use the Media Are Social Actors (MASA) paradigm as an overarching theoretical framework for examining the impact of social cues on user responses to media (Lombard & Xu, 2021). The MASA paradigm extends the Computers Are Social Actors (CASA) paradigm, which claims that people apply social rules in their interactions with computers (Nass & Moon, 2000), by considering a medium's social cues and the social signals these elicit to users as crucial for activating user responses. Lombard and Xu use Fiore et al.'s definition of social cues as "features salient to observers because of their potential as channels of useful information" (2013, p. 2). In contrast, social signals refer to the interpretation of a sender medium's social cues by the receiver (Fiore et al., 2013). Examples of a medium's social cues include gestures, motion, and language use. These can send out social signals of social identity, interactivity, and responsiveness to users (Lombard & Xu, 2021). This research focuses on cues signaling identity (agent type: human vs. chatbot) and responsiveness (i.e., the use or nonuse of verbal backchanneling cues).

A helpful model for investigating the impact of such cues is the Modality-Agency-Interactivity-Navigability (MAIN) model (Sundar, 2008). The MAIN model postulates that cognitive heuristics about an interaction's character and content are triggered when visual and

identity cues are used in the interface (Sundar, 2008). The features of an interface can thus shape users' interaction experience (Sundar, 2020). The model identifies four affordances, which are present in most media: Modality, Agency, Interactivity, and Navigability. In this research, we focus on agency, which refers to the information source whose identity is communicated by an interface, in our case, an agent, to the user (Sundar, 2008). Visual or verbal cues (e.g., a human-like picture or backchannel utterances) can communicate an interface's identity (Sundar, 2008). These cues can trigger cognitive heuristics, like the machine heuristic and the social presence heuristic, which likely affect users' responses to chat advisors (Sundar, 2008).

2.2.2 Agent Type: Chatbot vs. Human

According to the MAIN model, machines are expected to lack emotions and, thus, be objective, rule-governed, and invariant on the one hand (Sundar, 2008). Machine-like cues can positively influence credibility perceptions by triggering the machine heuristic (Sundar, 2008). For instance, Sundar and Nass (2001) found that news displayed by machine-like systems is perceived as more objective than news displayed by humans. On the other hand, machines are stereotyped as unemotional and cold (Sundar, 2020).

If the positive or negative impact of the machine heuristic prevails (i.e., whether people favor humans or machine agents) strongly depends on the task at hand (M. K. Lee, 2018). We argue that in developmental study advising, people might expect to talk to a human because they consider machines as “unfit for ‘human tasks’ that involve subjective judgments and emotional capabilities” (Sundar, 2020, p. 80). Advising prospective students about degree programs requires not only rational data processing but also intuitive judgments and interpersonal competencies (Mottarella et al., 2004). The outcome of developmental study advising can have decisive consequences for a person's life. Thus, in this context, people might prefer communicating with a human because they consider them more flexible, adaptable, and sympathetic than a machine—in our case, a chatbot. Previous studies found that machine-like cues hinder positive agent evaluations, while human-like cues promote positive assessments; for example, users rated human agents (vs. chatbots) higher regarding expected likability and social presence (Spence et al., 2014) and social attraction (Lew & Walther, 2023). Given theory and prior research, we expect participants to have a more positive attitude toward a human advisor. They should also consider a human more likable, intelligent, warm, and competent than the chatbot (Lou et al., 2022). Moreover, we expect perceived enjoyment, satisfaction, and use intention to be higher for the human advisor (Prahl & van Swol, 2021):

H1: Human identity cues have a positive effect on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use ($\mu_{\text{human identity cues}} > \mu_{\text{chatbot identity cues}}$).

2.2.3 Responsiveness

The general concept of responsiveness stems from the interpersonal relationship literature, where it refers to the likelihood of each partner responding to the other and the proportions of relevant and adequate responses (Davis, 1982). In close interpersonal relationships, responsiveness refers to “the processes through which relationship partners attend to and respond supportively to each other’s needs, wishes, concerns, and goals, thereby promoting each other’s welfare” (Reis & Clark, 2013, p. 400). It is considered pivotal for human attachment processes. Responsiveness depends on how partners perceive and respond to each other’s needs (Reis & Clark, 2013). When a partner feels that their needs are being met, feelings of closeness and mutual sympathy emerge. Partners who are responsive (i.e., psychologically empathetic, attentive, and supportive of one another) benefit in terms of liking, well-being, and satisfaction (Birnbaum et al., 2016; Davis & Perkowski, 1979; Reis & Clark, 2013). Responsive behavior manifests itself in asking questions, paralinguistic behavior in the form of backchannel utterances, summarizing and paraphrasing what has been said, and expressing understanding (Maisel et al., 2008). Backchannel cues signal “attention to, support or encouragement for, or even acceptance of the speaker’s message” (Mulac et al., 1998, p. 647). In service encounters, responsiveness has been shown to increase outcomes such as customer satisfaction and trust (De Ruyter & Wetzels, 2000). In human-machine communication (HMC), positive responsiveness effects on perceived competence, sociability, and willingness to use have been found for social robots (Birnbaum et al., 2016). There is first evidence that backchanneling increases the intention to use a chatbot (S. Lee et al., 2020). Taken together, we expect that responsive (vs. not responsive) advisors increase cognitive, affective, and behavioral user responses (Birnbaum et al., 2016; De Ruyter & Wetzels, 2000):

H2: Responsive verbal cues have a positive effect on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use ($\mu_{\text{responsive verbal cues}} > \mu_{\text{no responsive verbal cues}}$).

2.2.4 The Interplay Between Agent Type and Responsiveness

We do not know yet whether the effect of responsiveness is the same across both types of agents or not. According to expectation violation theory (Burgoon & Hale, 1988; Grimes et al., 2021), initial user expectations of an agent’s performance and whether they are confirmed or violated matter. Users attribute more attention to violated expectations than confirmed ones

(Burgoon & Hale, 1988). Expectation violations can be positive or negative: A positive violation is seen as beneficial (e.g., when the user perceives a conversational agent as better than expected). In contrast, a negative violation indicates that the user expected more of the agent than they received (Burgoon & Hale, 1988). By telling users they are about to chat with a chatbot, compared to a human, expectations of the agent are reduced (Grimes et al., 2021). When the chatbot employs responsive cues and thus strongly resembles the human agent in its conversational characteristics, a responsive chatbot might receive more positive user responses than a chatbot without responsive verbal cues. As overly anthropomorphic chatbots are often perceived as eerie (Mori et al., 2012), users might feel uncomfortable talking to a responsive chatbot. Therefore, one could also expect the responsive chatbot to receive more negative user responses than the chatbot without responsive verbal cues. Given these conflicting lines of reasoning and the lack of previous research (except, e.g., Beattie et al., 2020; Sundar et al., 2016), we formulate an exploratory research question:

RQ1: Is there an interaction effect between agent type (human vs. chatbot identity cues) and responsiveness (responsive vs. no responsive verbal cues) on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use?

2.2.5 Underlying Processes of Agent Type and Responsiveness Effects

To contribute to a more comprehensive understanding of the effects of agent type and responsiveness on user responses, we consider the processes underlying this relationship from three levels of analysis that relate to different aspects of interaction believed to be important in virtual interactions with social actors (Go & Sundar, 2019; Van der Goot & Etzrodt, 2023): the interaction in general, the dialogic nature, and the conversation content.

When looking at the interaction in general, we examine the mediating role of social presence. Social presence is defined as the perception of “being with another” (Biocca et al., 2003, p. 468). In HMC, it refers to the user’s perception of interacting with a social entity rather than a machine (Sundar, 2008). The concept has been shown to positively impact attitudinal and behavioral outcomes in virtual interactions (Gefen & Straub, 2004; Oh et al., 2018). Social presence is a fleeting judgment of an interaction influenced by the medium (Biocca et al., 2003). For instance, agents that provide human-like visual and verbal cues lead to stronger perceptions of social presence than agents that do not (S. Lee et al., 2020; Sundar, 2008). We thus expect human and responsive advisors to elicit higher levels of social presence, resulting in higher cognitive, affective, and behavioral user-related outcomes (Biocca et al., 2003; Go & Sundar, 2019).

H3: The effects of agent type on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use are mediated by social presence.

H4: The effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use are mediated by social presence.

Next, we look at the dialogic nature of the conversation. Dialogue is a fundamental feature of human conversations and provides the interlocutors with a sense of reciprocity, cooperation, and support (Kent & Taylor, 2002), also attributed to responsiveness (Reis & Clark, 2013). Conversations with responsively communicating agents should be perceived more as a dialogue than ones with an agent not using responsive verbal cues. Similar effects have been found for verbal cues signaling message contingency (Go & Sundar, 2019; Sundar et al., 2016). Perceived dialogue has been shown to increase advisor perceptions and usage intention (Go & Sundar, 2019). We argue that responsive verbal cues positively affect user responses via perceived dialogue:

H5: The effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use are mediated by perceived dialogue.

Closely related to the concept of perceived dialogue but focused more on the actual content of the conversation is the concept of feeling heard, defined as “the feeling that one’s communication is received with attention, empathy, respect, and in a spirit of mutual understanding” (Roos et al., 2023, p. 5). Responsive verbal cues could reinforce the user’s feeling of being heard, leading to more positive evaluations of the advisor and the interaction. As feeling heard is a new and under-researched concept, we want to answer the following research question:

RQ2: Does feeling heard mediate the effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction, and (h) intention to use?

All three potential mediating mechanisms are considered to play important roles in virtual interactions with social actors (Go & Sundar, 2019), which is why we believe they could operate in parallel (see Ischen, Araujo, van Noort, et al., 2020 for a similar approach). Feelings of social presence and being heard as well as the perception of dialogue in an interaction all involve notions of reciprocity, responsiveness, and mutual understanding (Roos et al., 2023; Sundar et al., 2016). Therefore, we will simultaneously investigate whether social presence,

perceived dialogue, and feeling heard mediate the responsiveness effects on our outcomes of interest. Entering all three mediators in the same model controls for shared variance and provides stronger evidence for conclusions about the underlying processes (Hayes, 2022).

2.2.6 Overview of the Current Studies

In three online vignette experiments conducted in 2021, we investigated the extent to which agent type and responsiveness influence users' responses to chat advisors. Vignette designs are common in HMC research (Abendschein et al., 2021; Beattie et al., 2020; Greussing et al., 2022). The local ethics committee of the Leibniz-Institut für Wissensmedien, Tübingen, approved the studies. Informed consent was obtained from participants before their participation. Preregistrations, materials, data, and additional results are freely accessible on OSF: <https://osf.io/w8dzv>. The deviations from the preregistrations are listed in Appendix A.

2.3 Study 1

2.3.1 Method

We conducted a 2 (agent type: chatbot vs. human) \times 2 (responsiveness: absence vs. presence of responsive verbal cues) between-subjects experiment. Participants were recruited via the online sampling platform Prolific. Of the 280 participants who completed the study, 253 passed the agent type manipulation check and were retained ($n_{\text{female}} = 101$, age: $M = 28.18$, $SD = 9.25$, range = 18–69) (power analysis in Appendix A). Participants were randomly assigned to one of four conditions. After providing informed consent, all participants saw a vignette in the form of a pre-recorded animated chat conversation between the study advisor, Sophie, and Marc, a prospective student at Sophie's university. The advisor asked the user several questions during the conversation to find out his interests. Based on Marc's answers, Sophie recommended suitable degree programs. Finally, we asked participants to complete a survey about their perceptions of the interaction and key demographics. The design of the chat interface resembled the design of contemporary messenger interfaces used in practice (Appendix A, Figure A.1). A robot icon was used to represent the chatbot, and it introduced itself as "Sophie, the chatbot of the student advisory service." The human advisor was represented with the portrait of a businesswoman, and she introduced herself as "Sophie, a student advisor" (see Go & Sundar, 2019 for a similar approach). All conversations were equal in content and without disruptions. Responsiveness was manipulated using short backchanneling cues and tokens like questions, paraphrases, and expressions signaling thinking processes (Maisel et al., 2008). Specifically, the responsive agents responded to the user's inputs with utterances like "Mhm,"

“Got it,” and “Hmm, let me think.” The responsive agent also asked the user for his name and repeated it in the following input (see stimulus conversations in Appendix A).

To measure the *attitude* toward the means of communication, we used a scale by Diers (2020). Specific cognitive user responses were assessed using the *likability* and *perceived intelligence* scales from the Godspeed Questionnaire (Bartneck et al., 2009) and the *warmth* and *competence* scales from Fiske (2018). Affective user response was assessed using a *perceived enjoyment* scale, and the behavioral aspect was reflected with an *intention to use* scale, both by Diers (2020). We asked participants to rate whether they would have found the advisor’s behavior satisfactory if they had been in the student’s position using Lagace et al.’s *satisfaction* scale (1991). We adopted the scales for *social presence*, *perceived dialogue*, and *feeling heard* from Gefen & Straub (2004), Roos et al. (2023), and Sundar et al. (2016), respectively. We included manipulation checks for agent type (adapted from Go & Sundar, 2019) and responsiveness (adapted from De Ruyter & Wetzels, 2000) to ensure effective manipulations. We assessed the same variables in all studies (see detailed measures in Appendix A, procedure in Table A.1, and descriptive statistics in Table A.2). Bivariate correlations were rather strong (Appendix A, Table A.3), ranging from $r = .39$ to $r = .89$ ($p < .001$). Attitude and intention to use, both drawn from the Technology Acceptance Model literature (Venkatesh, 2000) and relating to the communication with the organization, were strongly correlated ($r > .86$). Perceived dialogue and feeling heard also correlated strongly, potentially due to similarity in item content. Internal consistency of all constructs was satisfactory (Cronbach’s $\alpha > .80$).

2.3.2 Results

As intended, participants in the responsive conditions perceived the agent as significantly more responsive ($M = 5.89$, $SD = 1.03$) than those in the conditions without responsive verbal cues ($M = 5.15$, $SD = 1.36$), as a Welch two-sample t -test showed, $t(237.59) = 4.88$, $p < .001$, $d = 0.61$. To test H1 and H2 and to answer RQ1, we carried out a two-way multivariate analysis of variance (MANOVA). Using Pillai’s trace, there was a significant main effect of agent type on the outcomes, $V = 0.15$, $F(1, 248) = 5.40$, $p < .001$. Separate univariate analyses of variance (ANOVA) only revealed a significant agent type effect for satisfaction (g), $F(1, 248) = 6.72$, $p = .010$: Participants in the chatbot conditions tended to be more satisfied ($M = 5.56$, $SD = 1.24$) than participants in the human conditions (Satisfaction: $M = 5.13$, $SD = 1.48$), $t(239.74) = -2.51$, $p = .013$, $d = -0.32$. Neither the responsiveness effect, $V = 0.04$, $F(1, 248) = 1.16$, $p = .324$, nor the interaction between agent type and responsiveness, $V = 0.02$, $F(1, 248) = 0.74$, $p = .657$, were significant. We rejected H1 and H2. We did not perform mediation

analyses to test H3–H5 and to answer RQ2 because neither agent type nor responsiveness positively impacted the outcomes. We thus rejected H3–H5.

2.3.3 Discussion

Contrary to the hypotheses, neither human identity cues nor responsive behavior positively affected user responses. The zero effects of responsiveness are striking, given the significant responsiveness effects on the manipulation check. Moderation analyses did not yield significant results. The sample size was slightly below the target size due to our exclusion criterion. As we had based our power considerations on a small interaction effect that Go and Sundar (2019) found in a study where participants directly interacted with chat agents, the effect sizes in our vignette design could be even smaller. Hence, we decided to replicate our study with a larger sample.

2.4 Study 2

2.4.1 Method

The experimental design and measures were equal to those in Study 1. As we expected effects in Study 2 to be smaller than Go and Sundar’s interaction effect found in an experiment where users directly interacted with an agent, we determined a sample size of $N = 403$ a priori for a significant interaction effect of $f = .14$ ($\alpha = .05$, $1 - \beta = .80$). We aimed to recruit 403 participants via university mailing lists. A total of 520 participants completed the study. We excluded 118 participants because they failed the agent type manipulation check and one who admitted to not having answered the questionnaire reliably, which led to a final sample of $N = 401$ ($n_{\text{female}} = 287$, age: $M = 24.26$, $SD = 6.11$, range = 18–69).

2.4.2 Results

As intended, participants in the responsive conditions scored significantly higher on the responsiveness manipulation check ($M = 5.86$, $SD = 1.20$) than participants in the conditions without responsive verbal cues ($M = 4.67$, $SD = 1.54$) as a Welch two-sample t -test showed, $t(357.33) = 8.51$, $p < .001$, $d = 0.86$. We conducted a two-way MANOVA to test H1 and H2 and answer RQ1. Using Pillai’s trace, there was a significant main effect of agent type, $V = 0.06$, $F(1, 394) = 3.36$, $p = .001$. The responsiveness effect, $V = 0.03$, $F(1, 394) = 1.58$, $p = .129$ and the interaction between agent type and responsiveness, $V = 0.02$, $F(1, 394) = 0.95$, $p = .388$, were not significant in the multivariate model. Separate ANOVAs revealed a significant agent type effect for likability (b), $F(1, 394) = 4.92$, $p = .036$. Participants in the chatbot conditions rated the agent more likable ($M = 5.63$, $SD = 1.05$) than participants in the human conditions ($M = 5.41$, $SD = 1.07$), $t(397) = -2.21$, $p = .035$, $d = -0.21$. Although the responsiveness effect

was not significant, we computed separate ANOVAs, revealing positive responsiveness effects on warmth (d), $F(1, 394) = 6.32, p = .012$, and satisfaction (g), $F(1, 394) = 6.53, p = .011$. We rejected H1 and accepted H2d, g). We computed parallel multiple mediator models (Hayes, 2022) predicting warmth and satisfaction using the R package *lavaan* (Rosseel et al., 2021) but did not find significant indirect effects (Appendix A, Figures A.2, A.3).

2.4.3 Discussion

Like in Study 1, human identity cues did not significantly improve user responses in Study 2. However, we found significant effects of responsive conversational cues on warmth and satisfaction. We conducted a third study to clarify our findings from Studies 1 and 2.

2.5 Study 3

2.5.1 Method

We collected data from 418 participants via the crowdsourcing platform Clickworker. Three hundred fifty-one participants passed the agent type manipulation check and were retained ($n_{\text{female}} = 127$, age: $M = 38.53, SD = 12.28$, range = 18–73). The experimental design and measures were equal to those we used in Studies 1 and 2⁶.

2.5.2 Results

Participants in the responsive conditions scored significantly higher on the responsiveness manipulation check ($M = 6.08, SD = 0.88$) than participants in the conditions without responsive verbal cues ($M = 5.54, SD = 0.96$), as shown in a two-sample *t*-test, $t(349) = 5.54, p < .001, d = 0.59$. To test H1–H2 and to answer RQ1, a two-way MANOVA was carried out. Using Pillai's trace, there was a significant main effect of agent type, $V = 0.08, F(1, 347) = 4.11, p < .001$, on the outcome variables. The responsiveness effect, $V = 0.04, F(1, 347) = 1.66, p = .108$, and the interaction between agent type and responsiveness, $V = 0.02, F(1, 347) = 1.04, p = .403$, were not significant in the multivariate model. Separate ANOVAs only revealed a significant agent type effect on competence (e), $F(1, 347) = 12.49, p < .001$, and a marginally significant responsiveness effect on likability (b), $F(1, 347) = 3.85, p = .051$. Follow-up tests yielded that participants in the human conditions rated the agent more competent ($M = 5.76, SD = 0.93$) than participants in the chatbot conditions ($M = 5.36, SD = 1.12$), $t(349) = 3.54, p < .001, d = 0.01$. We, therefore, accepted H1e) and rejected H2. As we

⁶ Study 3 was designed to additionally explore the impact of agent response time on users' responses, so response time (immediate or dynamically delayed) was included as a third experimental factor. As the results were not decisive for answering the research question and for reasons of space, they are not part of the published article, but are presented in Chapter 2.9 of this dissertation.

did not find significant positive agent type or responsiveness effects (H2), we did not compute parallel multiple mediator models.

2.5.3 Discussion

Participants perceived the human agent as more competent than the chatbot, but no significant responsiveness effects were found. Still, a considerable proportion of participants did not recognize the alleged human, reducing the sample and resulting in a power loss for estimating interaction effects (posthoc power = 74.40%, $N = 351$, $\alpha = .05$, $f = .14$). To mitigate the potential power issues of Studies 1 and 3, we conducted analyses based on the pooled data from all studies.

2.6 Additional Analyses: Pooled Data

Using the pooled data ($N = 1,005$) and controlling for study number, we performed an exploratory MANCOVA to clarify the main effects on the outcomes. Using Pillai's trace, the inclusion of study number as a control variable indicated differences between studies, $V = 0.11$, $F(2, 995) = 7.19$, $p < .001$. Specifically, participants in Study 2 showed lower values on all outcomes than those in Studies 1 and 3, pointing to a generational effect. Study 2 comprised university students who were younger on average than participants in Studies 1 and 3 and, thus, may have had more experience with chat advisors. In contrast to the individual study results, using Pillai's trace, we found significant main effects of agent type (H1), $V = 0.07$, $F(1, 995) = 9.10$, $p < .001$, and responsiveness (H2), $V = 0.03$, $F(1, 995) = 3.25$, $p = .001$. The interaction between agent type and responsiveness (RQ1) was not significant, $V < .01$, $F(1, 995) = 0.59$, $p = .783$. Univariate ANCOVAs and pairwise comparisons yielded significant positive effects of human identity cues on competence (e) and intention to use (h). Participants perceived the interaction with the human agent as less enjoyable than the interaction with the chatbot. In addition, significant positive responsiveness effects emerged for likability (b), warmth (d), and satisfaction (g), strengthening the individual study findings (Table 1, with adjusted means in Table 2).

Table 1 Two-Way ANCOVA Statistics and Effect Sizes for Study Variables (Pooled Data)

Variable	Effect	<i>F</i> ratio	<i>p</i>	η^2_{partial}
Attitude	Study	15.80	< .001	.03
	AT	3.27	.071	.00
	R	0.64	.422	.00
	AT × R	0.44	.510	.00
Likability	Study	3.23	.040	.01
	AT	3.79	.052	.00
	R	6.12	.014	.01
	AT × R	0.10	.755	.00
Perceived intelligence	Study	16.43	< .001	.03
	AT	2.41	.121	.00
	R	1.97	.161	.00
	AT × R	0.34	.558	.00
Warmth	Study	10.45	< .001	.02
	AT	0.51	.475	.00
	R	8.93	.003	.01
	AT × R	0.71	.400	.00
Competence	Study	7.02	< .001	.01
	AT	7.32	.007	.01
	R	0.04	.847	.00
	AT × R	0.25	.616	.00
Perceived enjoyment	Study	26.27	< .001	.05
	AT	4.95	.026	.01
	R	2.67	.102	.00
	AT × R	0.01	.927	.00
Satisfaction	Study	14.29	< .001	.03
	AT	3.28	.070	.00
	R	6.09	.013	.01
	AT × R	1.11	.293	.00
Intention to use	Study	21.83	< .001	.04
	AT	6.90	.009	.01
	R	0.48	.487	.00
	AT × R	0.15	.694	.00

Note. $N = 1,005$. ANCOVA = analysis of covariance. Study = study number, AT = agent type, R = responsiveness. $df = 1,995$, except $df_{\text{Study}} = 2,995$.

Table 2 Adjusted Means and Effect Sizes for Study Variables (Pooled Data)

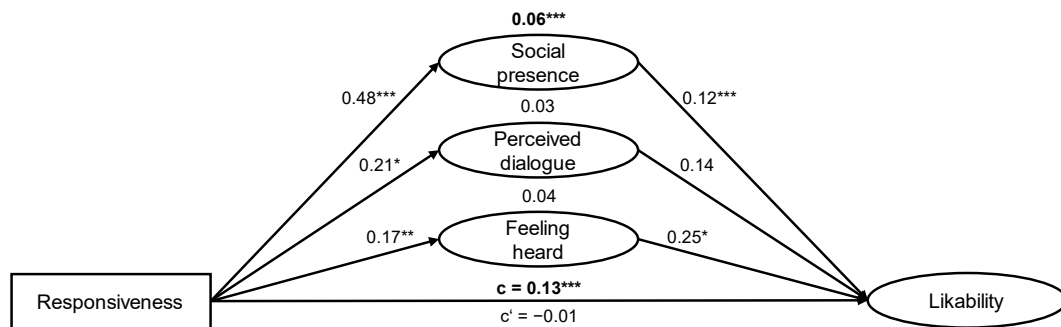
Variable	Agent Type				Responsiveness			
	Chatbot	Human	<i>p</i>	η^2_{partial}	Absent	Present	<i>p</i>	η^2_{partial}
Attitude	4.45	4.64	.068	.00	4.50	4.58	.435	.00
Likability	5.70	5.57	.050	.00	5.55	5.71	.013	.01
Perceived intelligence	5.58	5.68	.125	.00	5.58	5.67	.175	.00
Warmth	5.21	5.17	.481	.00	5.09	5.29	.003	.01
Competence	5.47	5.64	.006	.01	5.55	5.54	.844	.00
Perceived enjoyment	4.27	4.06	.029	.00	4.10	4.26	.089	.00
Satisfaction	5.20	5.04	.076	.00	5.01	5.24	.012	.01
Intention to use	4.35	4.65	.009	.01	4.45	4.52	.517	.00

Note. $N = 1,005$. One-way ANCOVAs controlled for study number.

To analyze the underlying relationship processes between agent type and responsiveness and the outcomes, we computed three parallel multiple mediator models predicting the outcomes significantly impacted by responsiveness; that is, likability (b), warmth (d), and satisfaction (g), using the R package *lavaan* (Rosseel et al., 2021). Including social presence, perceived dialogue, and feeling heard allowed us to model our three potential mediation levels simultaneously. We operationalized the latent constructs using reflective measurement models composed of the corresponding items.⁷ All standardized factor loadings were sufficiently strong ($\lambda > .50$) and significant ($p < .001$). Knowing the mediators to be strongly correlated, we specified their covariances. We controlled the models for study number. Model fit was acceptable (Westland, 2015). As expected, we found high correlations between social presence and perceived dialogue ($r = .67$ in the likability, $r = .66$ in the warmth and satisfaction models), social presence and feeling heard ($r = .61$), and perceived dialogue and feeling heard ($r = .95$) throughout the models. Figures 2–4 display the results of the mediation models predicting likability, warmth, and satisfaction.

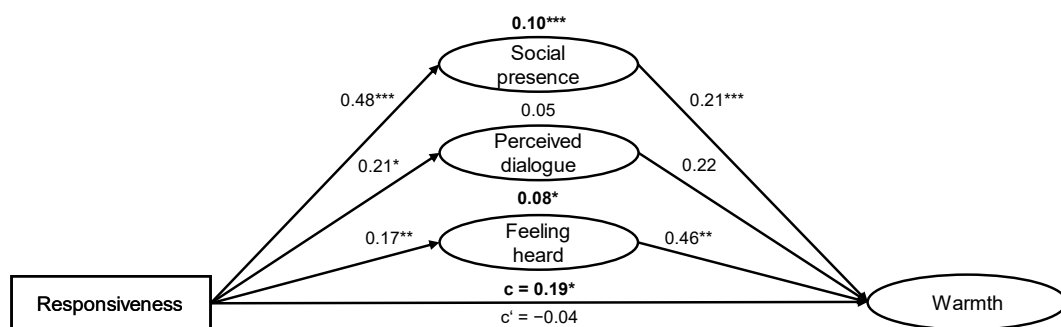
⁷ The standardized factor loadings of the two inversely coded feeling heard items were $< .50$. We thus followed Roos et al. (2023) in specifying the covariance between the residuals accounting for different response behaviors for inversely coded items. The error terms correlated moderately ($r = .42$, $p < .001$).

Figure 2 Parallel Multiple Mediator Model for the Effect of Responsiveness on Likability (Pooled Data)



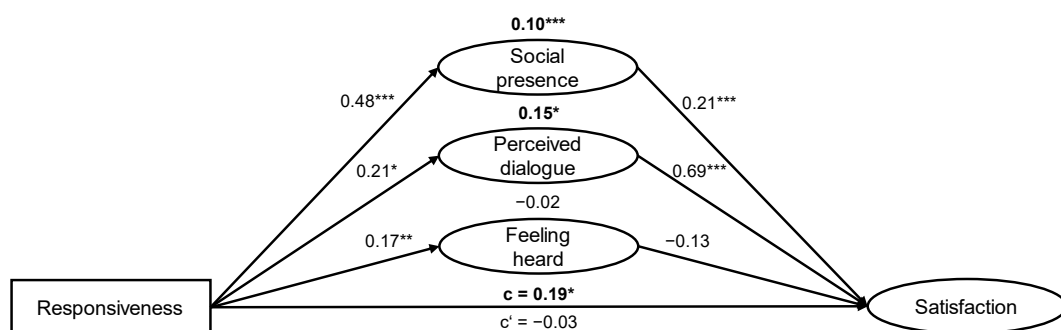
Note. $df = 261$. $\chi^2 = 1592.500$, $p < .001$, CFI = .914, RMSEA = .071, $CI_{RMSEA}(.068, .075)$, SRMR = .063. Unstandardized coefficients. Controlled for study number. $R^2_{Likability} = .473$, $R^2_{Social\ presence} = .054$, $R^2_{Perceived\ dialogue} = .014$, $R^2_{Feeling\ heard} = .009$.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 3 Parallel Multiple Mediator Model for the Effect of Responsiveness on Warmth (Pooled Data)



Note. $df = 285$. $\chi^2 = 2406.803$, $p < .001$, CFI = .875, RMSEA = .086, $CI_{RMSEA}(.083, .089)$, SRMR = .071. Unstandardized coefficients. Controlled for study number. $R^2_{Warmth} = .734$, $R^2_{Social\ presence} = .054$, $R^2_{Perceived\ dialogue} = .015$, $R^2_{Feeling\ heard} = .009$.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 4 Parallel Multiple Mediator Model for the Effect of Responsiveness on Satisfaction (Pooled Data)



Note. $df = 238$. $\chi^2 = 1552.062$, $p < .001$, CFI = .918, RMSEA = .074, $CI_{RMSEA}(.071, .078)$, SRMR = .063. Unstandardized coefficients. Controlled for study number. $R^2_{Satisfaction} = .592$, $R^2_{Social\ presence} = .054$, $R^2_{Perceived\ dialogue} = .015$, $R^2_{Feeling\ heard} = .010$.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Significant positive indirect effects emerged for likability via social presence, for warmth via social presence and feeling heard, and for satisfaction via social presence and perceived dialogue. The direct effects (i.e., the effects of responsiveness on the outcomes when the mediators were included in the model) were not significant, suggesting full mediations.

2.7 General Discussion

We conducted three vignette experiments to answer the question: To what extent do agent type and responsiveness influence users' responses to chat agents? Three key findings and subsequent implications emerged.

First, no study clearly supported our hypothesis on the positive impact of human identity cues on user responses (H1). The pooled analyses, however, suggested that participants considered the human agent more competent and indicated a higher intention to use it compared to the chatbot. This aligns with our hypothesis and suggests the machine heuristic does not hold for human tasks like developmental study advising (Sundar, 2020). Still, participants preferred the chatbot in terms of perceived enjoyment. There were no effects on the other outcomes, aligning with prior research suggesting little overall difference in the perception of humans and anthropomorphic chatbots (Beattie et al., 2020; Nass & Moon, 2000). Only the pooled analyses yielded small effects of human identity cues, although statistical power was high. The agents' error-free answers and suitable program recommendations to the user might be the reason for the small effects. Data were collected before ChatGPT was launched, so the high quality of the answers allegedly stemming from a chatbot might have been surprising for the participants. Considering the rapid improvement of generative AI, the question of whether and how small identity cues affect chatbot evaluations gets even more important. We can conclude that regardless of how well a chatbot performs and how much people enjoy it, human agents seem to be preferred as study advisors of a university.

Second, we found significant positive responsiveness effects on warmth and satisfaction in Study 2 (H2). The pooled analyses confirmed these findings and yielded an additional significant positive responsiveness effect on likability. We showed that a responsive communication style elicits positive responses in contexts where agents must provide support and understanding. Positive responses were elicited regarding the agent's social traits like likability, warmth, and satisfaction, corresponding to earlier findings from interpersonal communication research (Davis & Perrowitz, 1979). Responsive cues alone might be too subtle to tell whether people will want to be advised by an agent and not as decisive for perceptions relating to the successful completion of a task. Once high efficiency, the primary reason for

using chatbots (Følstad & Skjuve, 2019), is reached, these softer cues might become more important.

Third, perceptions of the agent mediated the relationship between responsiveness and likability, warmth, and satisfaction (H4–H5, RQ2). Social presence consistently mediated the effects on likability, warmth, and satisfaction, whereas perceived dialogue and feeling heard only mediated the effect on satisfaction and warmth, respectively. Overall, this is consistent with our theoretical assumption that responsive verbal cues from an interlocutor signal understanding and support, thereby leading to more positive user responses (Reis & Clark, 2013). However, because perceived dialogue and feeling heard were highly correlated and had similarities in item content (Appendix A), controlling for one construct eliminated the respective other's effect. This raises the question of whether the variables represent different constructs. We showed that verbal cues have the potential to make users feel more socially present and heard (Lombard & Xu, 2021). How the interaction is perceived appears to be more critical to perceptions of the advisor's social attributes and satisfaction than individual aspects (e.g., its dialogic nature and conversation content).

Interestingly, responsiveness did not interact with agent type. Responsive cues seem equally important to peoples' perception of chatbots and humans. Future research could employ other interindividual moderators that could affect the effect of agent type on user outcomes (e.g., affinity for technological interaction; Franke et al., 2019). Context-specific differences could also be explored; for example, responsive cues might matter more when a chatbot serves as an emotional support tool (e.g., Birnbaum et al., 2016 for social robots).

Agent type influenced certain outcomes, while responsiveness influenced others. The machine heuristic suggests machine actors are viewed as more objective, rule-based, and competent than humans (Sundar, 2008). Our results challenge this, as competence and intention to use were higher in the human conditions. The task of study advising, which we consider a *human task* at its core, might be the reason (Sundar, 2020). Additionally, the items used to assess intention to use referred to the way of communication with an organization. Thus, even if chatbots perform just as well as humans, a preference for talking to humans and an aversion to the use of automation and algorithms in universities' communication remain (Dietvorst et al., 2015).

We aimed to investigate users' perceptions of the agent and the interaction as well as the classical technology acceptance variables attitude and intention to use (Venkatesh, 2000). Agent and interaction perceptions are established antecedents of attitude and use intention, whereas the latter can predict actual usage (i.e., adoption; Diers, 2020; S. Lee & Choi, 2017;

Venkatesh, 2000). Responsiveness, in contrast, is more likely to impact social perceptions and seems relevant to users' satisfaction. For researchers more interested in the processes underlying chatbot adoption, the variables affected by responsiveness become relevant as they might mediate responsiveness effects on user satisfaction, which in turn might influence intention to use (Lou et al., 2022).

Our research contributes to the emerging research field of HMC regarding the impact of social cues on the perception and evaluation of machine agents (Gambino et al., 2020; Lombard & Xu, 2021). The different user responses to humans and chatbots suggest that not all social interactions with machine agents follow the media equation (i.e., not all machines are always perceived as social actors; Van der Goot & Etzrodt, 2023). There is reason to believe that users mindfully evaluate the source depending on situational factors (e.g., the interaction context) and dispositional factors (e.g., personality) (E.-J. Lee, 2024). Van der Goot and Etzrodt (2023) recommend conducting more qualitative research to unravel these processes and to understand "how users negotiate the blurring boundaries between humans and machines" (p. 27). This question will become more important as human-like chatbots based on generative AI continue to gain traction.

Previous research on study advising has shown that a warm and supportive advising style is an influential factor for satisfaction and is even more important than the advising approach (Mottarella et al., 2004). A warm communication style is often associated with developmental advising. Although prescriptive advising is task-oriented and focuses on explaining requirements and procedures, a more responsive style could improve student acceptance. We suggest that researchers investigating the differences between various advising approaches should pay more attention to the advisor's communication style, whether human or chatbot.

The results have implications for university practitioners considering using chatbots in developmental advising. While perceived competence and intention to use were higher for the human advisor, chatbot scores for these variables were above the scale means. So, chatbot support for student advising might be an efficient addition when financial resources or staff shortages are an issue. Leveraging the benefits of automated communication can thus be feasible without eliciting negative user responses. Yet, developers must ensure that the chatbots work well (e.g., by adequately exploiting the advantages of LLMs; Kasneci et al., 2023). But the way the chatbot presents the information is also critical. Study advising chatbots should be designed to evoke feelings of warmth and support, which have been shown to facilitate successful advising (Mottarella et al., 2004). Integrating responsive features into chat

interactions may help universities and schools build and maintain warm and supportive relationships with their (potential) students. A well-thought-out dialogue design can help integrate responsive verbal cues without too much financial or human effort.

2.7.1 Limitations and Future Research

Although vignette designs have high internal validity and give participants a unique perspective (Abendschein et al., 2021), they cannot offer as much ecological validity as experiments where participants directly interact with an agent. In our studies, participants were mere observers of the interaction, which could have increased their distance from the interaction, decreasing their involvement and identification with the user. The high nonrandom dropout rates due to failed agent type manipulation checks could have been related to the study design. To ensure experimental control, we kept the layout and content across all conditions constant. We thus manipulated agent type only in terms of the agent's introduction and avatar, which may have led participants to perceive the human agent as an anthropomorphic chatbot. Future studies could examine participants' direct interactions with chat agents to increase ecological validity. To ensure the comparability of our results, we used the same stimulus materials in all studies, which might have affected the validity of our results in case the stimuli did not optimally manipulate our independent variables. Future studies could use stimulus sampling (i.e., employ a variety of user-agent conversations) to reduce the impact of the unique features of a particular stimulus on the results and strengthen the conclusions (Jackson & Jacobs, 1983).

Our studies focused on investigating the effects of agent type and responsiveness on a wide range of dependent variables, including cognitive, affective, and behavioral user responses, but not on the relationships between the outcomes. As there is a plethora of scales from different disciplines that measure similar constructs (e.g., human-likeness perceptions; Ischen et al., 2023), researchers call for common conceptualizations and measurement scales for key outcomes (Følstad et al., 2021; Greussing et al., 2022). A systematic assessment and confirmatory factor analysis of common scales used in HMC research could shed light on what makes each construct unique, how the constructs are empirically related, and how they contribute to chatbot adoption.

2.8 Conclusion

In three experiments, we investigated the impact of agent type and responsiveness on a wide range of user-related outcomes in the context of study advising. Our results suggest that human agents are favored in terms of competence and intention to use but not in terms of perceived enjoyment. Further, the results indicate that responsiveness positively impacts users'

perceptions of agent likability, warmth, and satisfaction, mainly by increasing perceptions of the interaction. Our studies add novel insights to the literature on human-machine communication and offer two practical implications: First, our findings may encourage educational organizations to support their study advising departments with chatbots. Second, the use of responsive language by human agents and chatbots could help organizations build and maintain healthy and sustainable relationships with their (potential) students. Due to significant advances in generative AI, we can expect that people will increasingly be unable to distinguish whether they are interacting with a human or a chatbot in the future. Therefore, it will continue to be crucial to systematically investigate the role of relatively small social cues in the perception and evaluation of AI-based chatbots.

2.9 Additional Results: The Impact of Response Time (Study 3)

In this chapter, I report the results of additional analyses that are not included in the published article because they were not vital to answering the article's research question.

Response time, that is, the time the agent needs to send a reply to the user's message (Moon, 1999), is a conversational feature grounded in responsiveness (Van Pinxteren et al., 2020). It can be static (with or without delay) (Holtgraves et al., 2007), or dynamic, taking into account the complexity of the user's message and the agent's response (Gnewuch et al., 2018). Very short response times may suggest a lack of thought and cognitive effort in human-to-human interactions, while longer ones might induce a sense of deception (Ho et al., 2016; Moon, 1999). Users generally expect immediate responses from chatbots and browser-based systems which can increase perceptions of responsiveness and satisfaction (Holtgraves et al., 2007; Hoxmeier & DiCesare, 2000; Lew & Walther, 2023). Yet, delayed chatbot response times can also be beneficial as they may enhance social presence, intelligence and humanness perceptions and user satisfaction (Gnewuch et al., 2018; Kaptein et al., 2011). Because prior research on the effects of response times have come to conflicting conclusions (Gnewuch et al., 2018; Holtgraves et al., 2007), we wanted to answer the question: To what extent does agent response time (instant vs. dynamically delayed) influence (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction and (h) intention to use?

We acknowledge that instant response times are unrealistic for human agents as they require time to process the user's input and formulate and type a response. Thus, we expected valid response time effects mainly in the chatbot conditions. Dynamically (vs. instantly) responding chatbots could elicit more positive user responses due to their human-like behavior, or more negative user responses as users perceive them as eerie. In addition, we explore if response time and responsiveness interact to impact our outcomes. While responsive cues and dynamically delayed responses are expected to positively influence user responses, perceived inconsistencies might arise with instant but responsive or dynamic but non-responsive agents. Given limited previous research and conflicting theoretical arguments on the interactions between agent type, responsiveness, and response time, we aimed to explore how the factors interact: Are there interaction effects between agent type, responsiveness, and response time on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) satisfaction and (h) intention to use?

We differentiated between dynamically delayed and instant responses (Gnewuch et al., 2018). The agent responded one second after the user's input in the instant conditions.

Depending on input length, content complexity and agent response, the agent responded two to six seconds after the user's inputs in the dynamic conditions. User responses appeared dynamically delayed, but response times were equal in each condition.

Two-sample *t*-tests to answer RQ1 yielded no significant response time effects on either of the dependent variables (Table 3). A three-way ANOVA was computed for each outcome variable to answer RQ2 but yielded no significant interaction effects (Table 4).

The results contrast previous findings suggesting a positive influence of dynamic response times on outcomes such as satisfaction (Gnewuch et al., 2018). We could not directly compare the chatbot and the human conditions regarding response times due to potentially different expectations (Lew & Walther, 2023). Participants might have perceived the dynamically responding chatbot as relatively slow while perceiving the human advisor who responded dynamically as relatively fast. The lack of a significant effect could also be due to the response time manipulation. We employed a full factorial between-subjects vignette design to be able to compute all main and interaction effects. Therefore, we included scenarios that rarely (or never) occur (Auspurg et al., 2009), for example, the conditions in which the human agent responded instantly to the user's input. This could have irritated participants, as some of them let us know via the comment field in the questionnaire. Those conditions merely served as control groups whose data we did not interpret substantially. Further, we allowed participants to reread the chat conversation after the animation had ended to fully grasp the other manipulations (i.e., agent type and responsiveness). While scrolling through the conversation again, the response time manipulation could have been mitigated and not as salient as intended. Future research should expose participants to varying response times in real chatbot interactions.

Table 3 Means, Standard Deviations, and Test Statistics in Study Variables by Response Time (Study 3)

Variable	Instant		Dynamic		<i>t</i>	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Attitude	4.78	1.62	4.73	1.60	-0.27	.790	-0.03
Likability	5.70	0.93	5.74	1.00	0.32	.751	0.03
Perceived intelligence	5.63	0.95	5.69	1.04	0.60	.548	0.06
Warmth	5.26	1.11	5.32	1.12	0.49	.624	0.05
Competence	5.57	1.00	5.51	1.12	-0.57	.568	-0.06
Perceived enjoyment	4.57	1.43	4.53	1.57	-0.27	.788	-0.03
Satisfaction	5.25	1.37	5.35	1.42	0.70	.484	0.07
Intention to use	4.79	1.66	4.82	1.69	0.21	.836	0.02

Note. *df* = 349. *p*-values were adjusted using the Holm (1979) correction.

Table 4 Three-way ANOVA Statistics and Effect Sizes for Study Variables (Study 3)

Effect	Attitude		Likability		Perceived intelligence		Warmth		Competence		Perceived enjoyment		Intention to use		Satisfaction	
	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}
AT	2.13	.01	0.04	< .00	1.78	.01	3.13 [°]	.01	12.48 ***	.04	0.24	< .00	3.26 [°]	.01	0.78	< .00
R	0.32	< .00	3.83 [°]	.01	0.51	< .00	0.20	.01	0.13	< .00	0.45	< .00	0.17	< .00	0.57	< .00
RT	0.09	< .00	0.11	< .00	0.33	< .00	0.21	< .00	0.44	< .00	0.06	< .00	0.03	< .00	0.47	< .00
AT × R	0.32	< .00	0.29	< .00	0.01	< .00	0.10	< .00	0.24	< .00	1.36	< .00	< 0.00	< .00	0.28	< .00
AT × RT	0.22	< .00	0.35	< .00	0.17	< .00	0.18	< .00	0.09	< .00	0.27	< .00	0.06	< .00	2.14	.01
R × RT	0.96	< .00	1.55	.01	0.34	< .00	0.11	< .00	0.05	< .00	0.27	< .00	0.74	< .00	0.46	< .00
AT × R × RT	0.67	< .00	0.05	< .00	0.87	< .00	0.85	< .00	3.01 [°]	.01	0.69	< .00	0.38	< .00	1.62	.01

Note. $N = 351$. AT = Agent Type, R = Responsiveness, RT = Response Time. $df = 1, 343$.

[°] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

3. Free Text Interaction and Social Error Responses

Statement of Contribution

Stefanie Helene Klein developed the research question, created the stimulus materials, conducted data collection, management, and analyses, and drafted the first version of the manuscript. Dominik Papies and Sonja Utz supervised the research project, gave feedback on conceptualization and methodology, and edited and revised the manuscript. This chapter presents a slightly modified version of the published article:

Klein, S. H., Papies, D., & Utz, S. (2024). How interaction mechanism and error responses influence users' responses to customer service chatbots. *International Journal of Human-Computer Interaction*, 1–19. <https://doi.org/10.1080/10447318.2024.2351707>

Acknowledgments

We thank an anonymous firm for allowing us to use and adapt their chatbot implementation. We thank Dominik Warlich for his expertise and valuable support in developing the chatbot and Natalie Kraus and Anne Bucher for their assistance in preparing the stimulus materials.

Abstract

As more and more people regularly interact with chatbots in their everyday lives, it is crucial to understand how users perceive and evaluate them. This study examined whether free text (vs. button) interaction with and social (vs. neutral) error responses of chatbots effectively improve user responses. Using an online experiment ($N = 416$) in which participants interacted with a customer service chatbot, we investigated how interaction mechanism and error response influence perceived warmth and competence, satisfaction, usage intention, and emotional connection with the company. Contrary to our hypotheses, we found no evidence that free text interaction increases acceptance. Interacting via buttons was preferred, potentially because participants perceived the free text chatbot as less warm and competent. Interestingly, social error responses only increased satisfaction. We discuss the results considering recent advances in Large Language Models and highlight how the findings can guide further research and help practitioners develop efficient, user-friendly chatbots.

Keywords: chatbot, interaction mechanism, error response, customer service, satisfaction, intention to use, large language model

3.1 Introduction

Even before ChatGPT (OpenAI, 2023) and Large Language Models (LLM) started to dominate the headlines around the globe in 2023, improvements in artificial intelligence (AI) and automation drove the proliferation of chatbots in various sectors and will continue to do so in the future. Implementing chatbot technology enables companies to save staff and training costs by automating customer service processes like recommending products, managing complaints, and answering questions (Desai, 2023; Rivas & Zhao, 2023). According to a recent study by Gartner, about a quarter of companies will adopt chatbots as their main channel for customer service by 2027 (Costello & LoDolce, 2022). Accordingly, more and more people will regularly interact with chatbots in their everyday lives, which renders it crucial to systematically study how people perceive and evaluate them. For companies, implementing chatbots can lead to cost savings, but requires first investments in skilled personnel and technology (Vial et al., 2023). Our study can provide guidance to companies and practitioners on how to design efficient, accurate, and accessible chatbots that can improve the customer's service experience while being economically beneficial to the company.

Chatbots are automated text-based dialogue systems with which humans can interact via natural language (Shawar & Atwell, 2007). One can communicate with chatbots and not only through them, which makes them communicative subjects (Guzman, 2018). Many customer service chatbots rely on AI-based technologies such as natural language recognition and processing. Although labeled the gold standard in human-chatbot interaction, free text interactions with chatbots often result in frustration and abandoned conversations because the chatbot does not understand user queries, phrases, or context (Klopfenstein et al., 2017). In addition, free text interactions require users to input text through typing on their keyboards, which can be a cumbersome and unattractive task, particularly for users who are not experienced at typing. Structuring elements, such as buttons that display pre-defined answer options for users to click on, are already employed in state-of-the-art chatbots as well as LLM-powered chatbots, e.g., Microsoft's Bing chatbot (Bing, 2023). Buttons make the conversation flow more transparent but can also negatively affect perceived humanness, social presence, and user-chatbot bonding (Diederich et al., 2019; Mai et al., 2022). In this study, we thus want to investigate how the interaction mechanism (buttons vs. free text) affects the perception of the chatbot.

Most people use chatbots for productivity reasons, i.e., because they are easy to use and respond to requests quickly (Brandtzaeg & Følstad, 2017). However, chatbots that do not meet these requirements might negatively impact customers' trust and word-of-mouth (Seeger &

Heinzl, 2021). Since non-understandings are common in user-chatbot conversations, it is crucial to prevent users from losing motivation and abandoning the conversation (Ashktorab et al., 2019; Følstad & Taylor, 2020). How the error response is presented, i.e., whether the chatbot apologizes and reveals its weaknesses, is a potential factor in improving user perception but has been neglected in research on chatbots so far (Benner et al., 2021). We, therefore, want to study how a social response to errors affects users' perceptions of the chatbot, satisfaction, use intention, and emotional connection with the company.

The aim of our paper is to answer the following research question: How do free text interaction and social error responses influence perceived warmth and competence, satisfaction, intention to use, and emotional connection with the company? We use the Media Are Social Actors (MASA) paradigm (Lombard & Xu, 2021), which extends the Computers Are Social Actors (CASA) paradigm (Nass et al., 1994), social response theory (Lombard & Xu, 2021; Nass & Moon, 2000) and the concept of anthropomorphism (Epley et al., 2007) to derive our hypotheses. We test our hypotheses using an online experiment where participants interact with a state-of-the-art chatbot in a typical online retail customer service scenario.

Previous studies have investigated either the effects of a chatbot's interaction mechanism (Diederich et al., 2019; Haugeland et al., 2022) or the effects of a chatbot's verbal cues (Chattaraman et al., 2019; S. Lee et al., 2020) without considering possible interaction effects between the two. We complement existing research on the impact of combinations of different types of social cues on user responses (Lombard & Xu, 2021) by examining whether interaction mechanism and a chatbot's reaction to errors interact in their effects on the outcomes of interest. In addition to the theoretical implications, studying this interaction may also have important practical implications for the design and implementation of human-like and efficient customer service chatbots by practitioners and companies. By analyzing whether two dimensions central to the evaluation of new agents, i.e., warmth and competence, mediate the effects of interaction mechanism and error responses on the outcome variables, we expand existing research by shedding light on the mechanisms underlying the relationship between social cues and user responses.

This study extends previous research on the impact of a chatbot's interaction mechanism (Diederich et al., 2019; Haugeland et al., 2022) and error handling (Ashktorab et al., 2019; M. K. Lee et al., 2010) by investigating their influence on five user-related outcomes. The five outcomes represent central variables for assessing user perceptions, user experiences, and the company perspective (Cheng & Jiang, 2020; Fiske et al., 2007; Verma et al., 2016). First, we look at the warmth and competence of the chatbot as perceived by users. Warmth and

competence are the basic dimensions along which people evaluate new actors (Fiske et al., 2007). Based on prior research on AI-based agents (Blut et al., 2021; Cameron et al., 2021), we argue that more human-like interactions with a chatbot might be positively correlated with users' evaluations thereof. We do not include perceptions of agents that have been studied extensively, such as social presence or trust (Zierau et al., 2020). Negative emotions like anger (Crollic et al., 2022) are not considered either because the interaction in this study is unlikely to elicit much anger due to the experimental design, i.e., all participants have to solve the same fictitious tasks with the help of the chatbot. Second, we examine two central variables related to users' experiences with the system: satisfaction, i.e., the extent to which the chatbot experience meets their needs and expectations (ISO, 2018), and intention to use the system, i.e., the desire to use it in the future (Gopinath & Kasilingam, 2023). Satisfaction with a system is considered one of the main objectives of system design and evaluation in human-machine interaction (ISO, 2018). Both variables are commonly used to investigate user responses to machine agents in human-computer interaction research (Cheng & Jiang, 2020; De Andrés-Sánchez & Gené-Albesa, 2024). Third, we consider a company-related outcome, namely users' emotional attachment to it, i.e., the affinity between consumers and the company. Previous research has shown that anthropomorphic chatbot cues positively impact the relationship-building of companies with their customers (Araujo, 2018), which in turn has been shown to increase word-of-mouth, customer loyalty, and continuity expectation (Verma et al., 2016).

3.2 Theoretical Framework

3.2.1 Social Responses to Chatbots

An established framework for researching human-machine communication is the Computers Are Social Actors (CASA) paradigm. The CASA paradigm was derived from the media equation, which claims that humans treat computers and other media as real people (Reeves & Nass, 1996). According to the CASA paradigm, humans apply similar rules in their interactions with machines as in their interactions with other humans (Nass et al., 1994; Nass & Moon, 2000). The idea of humans treating machines as social actors is closely linked to the concept of anthropomorphism, that is, the tendency of humans to ascribe human characteristics, motivations, intentions, and emotions to non-human entities (Epley et al., 2007). Nass and Moon (2000) propose three main characteristics incorporated in machines leading people to act socially towards them. These properties are especially evident in text-based chatbots. The first characteristic refers to chatbots outputting written or spoken words. Second, chatbots are based on dialogue and turn-taking and are thus highly interactive. Third, chatbots fill roles traditionally occupied by humans, e.g., customer service agents, learning tutors, or mental

health advisors (Følstad et al., 2019). These characteristics make chatbots media agents, defined as “technological artifacts that demonstrate sufficient social cues to indicate the potential to be a source of social interaction” (Gambino et al., 2020, p. 73).

The Media Are Social Actors (MASA) paradigm structurally extends the CASA paradigm by emphasizing the impact of social cues on users’ responses to various kinds of media agents like social robots and voice- or text-based conversational agents (Lombard & Xu, 2021). Media agents provide verbal and non-verbal social cues that send out social signals, which in turn elicit social user responses (Feine et al., 2019; Go & Sundar, 2019; Lombard & Xu, 2021). Social cues are defined as “features salient to observers because of their potential as channels of useful information” (Fiore et al., 2013, p. 2). In contrast, social signals refer to how the perceiving person interprets the social cues of the sender media agent (Fiore et al., 2013). Examples of social cues include eye gaze, gesture, body movement, and language use, which can send out social signals like responsiveness, interactivity, and even personality and social identity (Feine et al., 2019; Lombard & Xu, 2021).

Social cues in media agents have previously been categorized in different ways. For example, Seeger et al. (2021) distinguish between human identity, verbal, and non-verbal cues, whereas Feine et al. (2019)’s taxonomy comprises verbal, visual, auditory, and invisible cues. This study investigates the effect of a chatbot’s verbal cues in the form of its responses to errors on user responses. Verbal cues are expressed in written or spoken words, which can signal responsiveness, attentiveness, understanding, support, and human-likeness (Feine et al., 2019; Lombard & Xu, 2021). Previous research has shown that we are more easily persuaded when we perceive a conversational partner to be similar vs. not similar in how we communicate verbally (Fogg, 2003). Even slight verbal variations, e.g., regarding the conversational style of the chatbot or short text messages, e.g., “I am sorry,” can convey human-likeness and, thus, influence our perceptions and evaluations of media agents (Chattaraman et al., 2019; J. Chen et al., 2023; Van Pinxteren et al., 2020). However, as our study not only examines verbal error responses but also different mechanisms of interaction between user and chatbot, we also make use of Fogg (2003)’s category of social dynamics cues. Social dynamics cues are “unwritten rules of how people interact with each other” (Fogg, 2003, p. 105) and entail turn-taking, cooperation, and reciprocity (Lombard & Xu, 2021). Cues representing social dynamics follow from the mechanism through which the interaction with a media agent takes place, e.g., through free text input or by clicking on buttons with pre-defined answer options (Jain et al., 2018). Social dynamics cues can thus influence the perceived engagement and interactivity of a media agent, which in turn can affect user responses (Haugeland et al., 2022). In the following two

sections, we lay out the current state of the research and present our hypotheses on the impact of interaction mechanisms and error responses on user responses.

3.2.2 Interaction Mechanism and User Responses to Chatbots

A factor that might strongly influence user acceptance is the interaction mechanism (Haugeland et al., 2022), also referred to as message form (Klopfenstein et al., 2017), interaction paradigm (Valério et al., 2020), interaction method (Mai et al., 2022) or turn-taking mechanism (Zierau et al., 2020). The interaction mechanism of a system refers to the kind of input it receives from the user during the interaction (Jaimes & Sebe, 2007). Different types of inputs are made in different ways: In this article, we distinguish between free text interaction, i.e., interaction with a chatbot by entering text via the keyboard, and button interaction, i.e., interaction with a chatbot by clicking on buttons with pre-defined answer options using the mouse.

Previous research on the effects of chatbots' interaction mechanisms on user responses has yielded mixed results. On the one hand, there is empirical evidence that using buttons reduces a chatbot's perceived humanness, potentially leading to lower user acceptance. In Diederich et al. (2019)'s experiment, for example, participants interacted with one of two chatbots where one was based on free text while the other was based on a combination of free text and preset answer buttons. The authors found that buttons reduced the perceived humanness and social presence of the conversational agent in the interaction. Along these lines, the experimental study by Mai et al. (2022) showed that free-text interaction improves the working alliance by enhancing the feeling of mutual respect, perceived sympathy, and appreciation by a coaching chatbot. A mixed-methods case study by Valério et al. (2020) showed that people with scientific backgrounds do not prefer free text over button chatbots. In turn, data from daily chatbot users analyzed by Jain et al. (2018) revealed that users preferred chatbots "that provided either a "human-like" natural language conversation ability or an engaging experience that exploited the benefits of the familiar turn-based messaging interface" (p. 895).

On the other hand, structured chatbot designs utilizing buttons might enhance user experience (Narducci et al., 2020). In an experiment on voting advice applications, Kamoen et al. (2022) found that buttons appeared more playful to all participants and easier to use to lower-educated participants than text fields. A randomized within-subjects experiment by Haugeland et al. (2022) found that button interaction leads to a higher pragmatic and hedonic user experience than free text interaction.

Based on the MASA paradigm and social response theory, we argue that users perceive interactions via free text as more human-like, i.e., more like their communication with other

humans, and, accordingly, to have more similar social dynamics than interactions via buttons. Thus, users perceive chatbots based on free text input as warmer. Based on higher similarity to human-to-human interactions, we also propose that chatbots based primarily on free text input are perceived as more competent as they seem able to understand the requests written by the user. In addition, free text interaction mimics natural human communication more adequately than button interaction, leading to users expending less cognitive effort (Gambino et al., 2020) and, in turn, to increased satisfaction, intention to use it, and emotional connection with the company.

H1. When users interact with a chatbot primarily via free text (as opposed to buttons), (a) perceived warmth and (b) competence of, (c) satisfaction with, (d) intention to use the chatbot, and (e) emotional connection with the company will be higher.

3.2.3 Responses to Errors and User Acceptance

As conversational breakdowns in user-chatbot conversations occur frequently, users should be prevented from losing motivation and abandoning the conversation (Ashktorab et al., 2019; Følstad & Taylor, 2020). Investigating the error repair strategies of chatbots is thus crucial. Repair strategies refer to strategies “that a chatbot [...] could adopt to tackle the above problems—providing evidence for the breakdown and supporting repair towards a desirable direction for the system model” (Ashktorab et al., 2019, p. 2).

Ashktorab et al. (2019) found proactive repair strategies, like providing options and explanations, were favored among several chatbot repair strategies. Sheehan et al. (2020) similarly found that a chatbot that tries to clarify its mistake increases people’s adoption intent compared to a chatbot that ignores its mistake. Positive effects of social error-handling strategies, such as offering an apology after an error to mitigate its negative impact, have also been demonstrated in social robots (Cameron et al., 2021; M. K. Lee et al., 2010). Besides apologizing for mistakes, social error-handling strategies include justifying or explaining the reasons for the error, promising to do better in the future (Robinette et al., 2015), as well as self-blame (Groom et al., 2010).

In particular, combinations of different error management strategies can lead users to continue the conversation (Benner et al., 2021). The combination of apologizing to the user for the error and admitting weaknesses has great potential to prevent conversation dropouts but has not been systematically studied for chatbots. We define an error response that contains verbal human-like cues like apologizing and admitting weaknesses as a social error response. A neutral response to errors does not contain verbal social cues. We argue that a social error response is

more efficient in increasing our outcomes of interest than a neutral error response because people perceive it to be more human-like. We propose the following hypothesis:

H2. When a chatbot's responses to errors are social (as opposed to neutral), (a) perceived warmth and (b) competence of, (c) satisfaction with, (d) intention to use the chatbot, and (e) emotional connection with the company will be higher.

3.2.4 The Interplay of Interaction Mechanism and Error Response

In line with the MASA paradigm and social response theory, social error responses in combination with free text (vs. button) input might make the chatbot appear more human-like as more social cues lead to more positive user responses (Lombard & Xu, 2021). Chatbots based on pre-defined buttons are perceived as more rule-based and less error-prone due to their structured design (Jain et al., 2018; Valério et al., 2020). However, suppose a button-based chatbot does not understand a request. In that case, users might expect a neutral and formal response rather than a social one, which might be more likely perceived as inconsistent, resulting in feelings of irritation and confusion among users. User responses might thus be less positive, as research on combining text-based chatbots with a human-like avatar (Ciechanowski et al., 2019) has shown. We do not expect substantial differences in user responses between free text and buttons when the error response is neutral since neutral error responses do not contribute as much to a human-like perception of the chatbot as social ones. Therefore, we hypothesize the following:

H3. The type of interaction mechanism moderates the effect of the error response on the outcomes, such that the effect of social (as opposed to neutral) error responses is stronger when the interaction mechanism is free text (as opposed to buttons).

3.2.5 Underlying Processes

When people are confronted with a new actor, they evaluate them on two basic dimensions: perceived warmth and competence (Fiske, 2018). According to the MASA paradigm and Social Response Theory, people treat technological entities, including chatbots, as social actors (Nass & Moon, 2000). We, therefore, argue that people evaluate a chatbot they encounter for the first time similarly to another person. An actor's first impression, e.g., regarding their likability or perceived intelligence, is often the most influential factor for their further assessment—and if positive, can lead to more positive evaluations (Bartneck et al., 2009). Previous research has shown that the more human-like, and therefore, similar to oneself, a machine is perceived, the more people tend to like and be socially attracted to it, and the more intelligent people perceive it (Beattie et al., 2020; Ciechanowski et al., 2019; Toader et al., 2020). Consequently, when people perceive a machine to be likable and intelligent, user

responses like the intention to use the machine turn out to be more positive (Blut et al., 2021; Cameron et al., 2021). Perceived intelligence refers to the extent to which people believe a machine is capable of independent thought, learning, and problem-solving; likability refers to qualities that elicit a favorable (first) impression (Bartneck et al., 2009). Free text interaction and social error responses are social cues that make the conversation with the chatbot more human-like. As H1 and H2 suggest positive effects of free text interaction and social error responses on satisfaction with, intention to use the chatbot, and emotional connection with the company, we assume these effects are mediated by perceived warmth and competence.

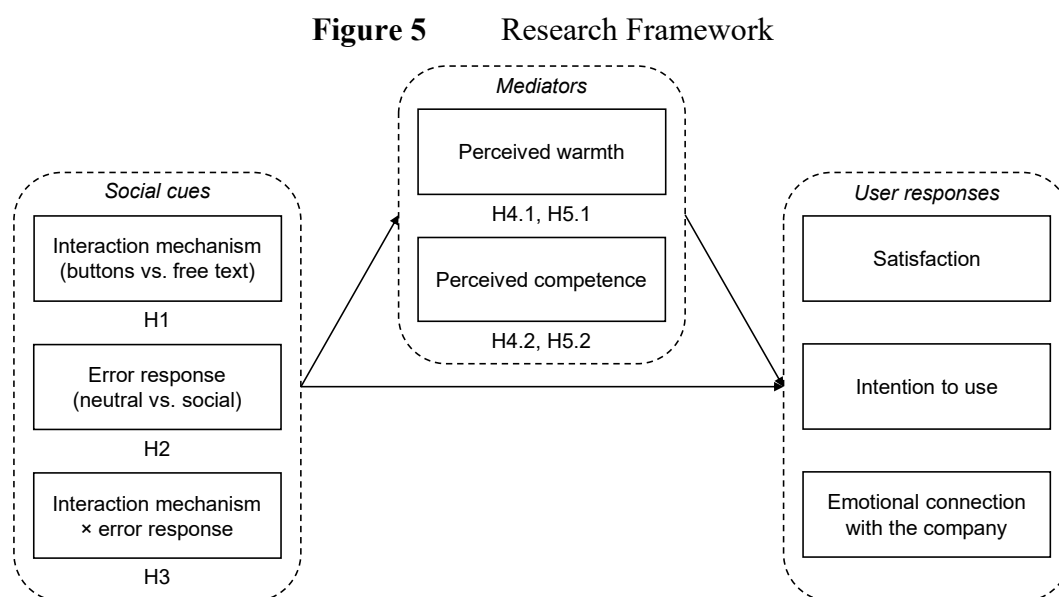
H4.1. Perceived warmth mediates the positive effects of free text interaction on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

H4.2. Perceived competence mediates the positive effects of free text interaction on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

H5.1. Perceived warmth mediates the positive effects of social responses to errors on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

H5.2. Perceived competence mediates the positive effects of social responses to errors on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

The research framework of this study is shown in Figure 5.



3.3 Method

3.3.1 Overview

We conducted a 2 (interaction mechanism: buttons vs. free text) x 2 (error response: neutral vs. social) between-subjects online experiment. After providing informed consent, participants were randomly assigned to one of four conditions where they interacted with a real customer service chatbot that we adapted to this specific research setting in collaboration with a company specializing in Conversational AI to make the setting as realistic as possible. We chose this context as chatbots are already frequently employed in customer service and continue to hold great potential in this area (Costello & LoDolce, 2022). Each condition comprised between 93 and 110 participants. The study was approved by the ethics committee of the Leibniz-Institut für Wissensmedien Tübingen (LEK 2022/046) and preregistered at <https://osf.io/shb6z>. Deviations from the preregistration can be found in Appendix B.

3.3.2 Participants

An a-priori power analysis to determine the minimum sample size to detect an effect of $f = 0.14$, with $\alpha = .05$ and a power ($1 - \beta$) of 80% yielded 403 participants (Faul et al., 2007). To account for potential exclusions, we recruited 450 German participants via the online sampling platform Prolific in October 2022. Respondents received £2.85 for participating in the experiment, which took around 20 minutes.

Following our preregistration, we excluded participants who had not followed the experimental instructions ($n = 30$) or experienced technical difficulties and could not interact with the chatbot ($n = 4$). The final sample comprised 416 cases. Participants' average age was 29.90 years ($SD = 9.16$, range = 18,71), and 50.00% identified as females. 61.78% indicated having used a chatbot at least once in the last three months.

3.3.3 Procedure

After starting the questionnaire, participants were asked to put themselves in a typical customer service scenario. They should imagine that they have placed an order and now have some questions to be answered and tasks to be completed by the chatbot (see scenario and instructions in Appendix B). All participants received the same set of tasks, i.e., they were to find out (1) the balance of the gift card used to pay, (2) what happens if the recipient is not at home when the package is delivered, (3) the current location of the package, (4) the cost of return shipping, and they were (5) to subscribe to the newsletter (optional). To ensure anonymous data collection, all participants were given the same fake customer credentials, such as an email address, order number, parcel tracking id, and pin code, to share with the chatbot

while working through the list of questions and tasks. We gave participants some general instructions for chatbot interaction: type their requests as free text in the text box if no options were offered, try to rephrase their request if the chatbot did not understand it, and type “I want to start over” if they wanted to start a new conversation. After reading the instructions, participants were prompted to click a button, opening a chat window embedded in the online survey. To make it easier for participants to answer the questions, they could see the list of tasks and the credentials while interacting with the chatbot. Participants received instructions to interact with the chatbot for a maximum of five minutes. To make the conversation more realistic and to be able to manipulate the experimental factor error response later, participants experienced at least one situation where the chatbot did not understand the user’s request during the interaction. We chose this type of error because failure to understand user requests is one of the most common mistakes users encounter when interacting with a chatbot (De Sá Siqueira et al., 2023). After the interaction, participants answered a questionnaire on their perceptions of the chatbot and the interaction. Participants could then comment on the study and were redirected to Prolific.

3.3.4 Stimuli and Manipulation of Independent Variables

To give our research practical relevance and increase external validity, we created the chatbots in collaboration with a company specialized in conversational AI. The partner provided access to their low-code Conversational AI platform, which allowed us to create four chatbot versions for experimental purposes. The chatbots were based on a set of technologies that power automated human-like interactions, e.g., natural language processing (NLP), which enables the chatbots to understand human language text, and the deep learning architecture BERT⁸ which represents the key functionality of the matching process between the intent predicted from a user’s chat message and the chatbot’s response. The conversations with the chatbots were largely automated, except for the error situations, which we predefined (Appendix B, Figure B.1). Specifically, when participants tried to solve tasks 2 (what happens if the recipient is not at home when the package is delivered) and 4 (the cost of return shipping), they were automatically redirected to separate conversational flows. No matter what they entered here, the chatbot did not understand their request and asked them to rephrase their

⁸ Bidirectional Encoder Representations from Transformers (BERT) is a language representation model developed by Google. BERT is used to pre-train bidirectional representations from unlabeled text. It can then be fine-tuned to create models for many language understanding tasks, e.g., question-answering and translation (Devlin et al., 2019).

question. After participants had rephrased, the chatbot provided the correct answer. As the interaction with the chatbot took place freely and unsupervised, some participants might have encountered more than two errors.

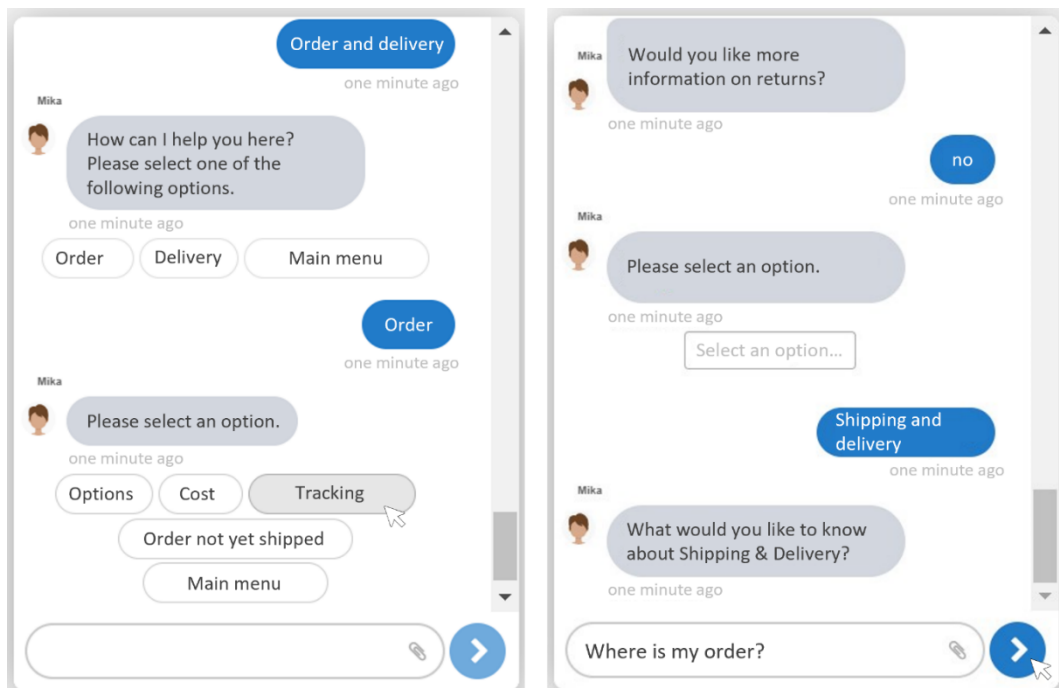
Anthropomorphic cues improve user responses to text-based chatbots, e.g., emotional connection with the company (Araujo, 2018) and user compliance (Adam et al., 2021). We thus equipped our chatbots with anthropomorphic cues, i.e., a gender-neutral name (“Mika”) and profile picture. The chatbots resembled currently employed customer service chatbots in design and functionality (Clustaar, 2019). The chatbots greeted the participants in all conditions, saying, “*Hello! I am Mika, the chatbot of myclothes.deu*”. Prior research has shown that dynamically delayed response times increase user satisfaction (Gnewuch et al., 2022). Thus, the chatbot typically responded after two seconds and after three seconds if its response was longer than two sentences or if the chatbot had a task to complete, such as registering the user for the newsletter.

Interaction Mechanism. Interaction mechanism was manipulated by creating two chatbot versions: One chatbot required participants to enter their requests via their keyboard into the text field. The other chatbot was based on pre-defined buttons, i.e., participants had to click the respective button to make their requests. No free text input was possible when participants could select a button. The chatbots were based either mainly on free text or mainly on buttons. Since there is no pure type in the real world, the free text chatbot also had a button menu at the beginning, and the button chatbot also had two free text input situations where participants had to enter the fake credentials we provided, such as their gift card pin code or newsletter signup email address.

To test the chatbots’ functionality, we conducted a two-group (button vs. free text chatbot) pre-study with 50 subjects recruited via Prolific ($n_{\text{Female}} = 24$, $M_{\text{Age}} = 29.6$, $SD_{\text{Age}} = 10.20$). In addition to common usability scales, such as the User Experience Questionnaire (UEQ; Schrepp et al., 2017), the System Usability Scale (SUS; Bangor et al., 2008), and the Chatbot Usability Questionnaire (CUQ; Holmes et al., 2019), participants were asked open-ended feedback questions about positive and negative aspects of the chatbot as well as selected demographic information (Appendix B). The pretest showed that both chatbots worked reliably. Usability scores were in the acceptable range. UEQ was measured on a scale from -3 to $+3$ where values between -0.8 and 0.8 represent a neutral evaluation of the corresponding scale, values above 0.8 a positive, and values below -0.8 a negative evaluation (Schrepp et al., 2017): $M_{\text{Button}} = 1.06$, $SD_{\text{Button}} = 0.86$, $M_{\text{Free text}} = 0.67$, $SD_{\text{Free text}} = 0.85$. A two-sample t -test yielded no significant differences between the two conditions ($t(46.95) = 1.59$, $p = .118$). We assessed SUS

and CUQ on scales from 1 = *fully disagree* to 5 = *fully agree* and converted the results to an overall score (0–100). According to the interpretation scheme (Bangor et al., 2008), both chatbots yielded acceptable SUS scores ($M_{\text{Button}} = 83.20$, $SD_{\text{Button}} = 11.10$, $M_{\text{Free text}} = 75.25$, $SD_{\text{Free text}} = 12.00$) with the button chatbot being slightly more acceptable, as shown in a two-sample t -test ($t(46.32) = 3.40$, $p = .020$). CUQ scores were similar in magnitude and direction ($M_{\text{Button}} = 76.75$, $SD_{\text{Button}} = 11.90$, $M_{\text{Free text}} = 69.16$, $SD_{\text{Free text}} = 10.1$, $t(46.32) = 2.41$, $p = .020$). Based on the participants' comments, we made minor improvements to both chatbots, e.g., regarding knowledge base and category naming. Figure 6 shows exemplary screenshots of the button and free text chatbots used in the experiment.

Figure 6 Translated Exemplary Screenshots of Button (left) and Free Text Chatbots



Error Response. After ensuring that the button and free text chatbots functioned reliably and that each conversational flow included two situations where the chatbot did not understand the user's request, we focused on manipulating the chatbots' error response, our second experimental factor. In the social error response conditions, the chatbot responded to a mistake it made socially, i.e., its response comprised an apology and a disclosure statement that it is still learning: *"I am sorry, unfortunately, I did not understand that. I am still learning. Please rephrase your request"*. In the neutral error response conditions, the chatbot responded to errors neutrally, i.e., it simply said: *"I did not understand that. Please rephrase your request"*. We provide translated exemplary conversations for each condition in Appendix B, Figure B.1. Videos of conversations in the original language are available in our OSF repository: <https://osf.io/j8hde/>.

3.3.5 Measures

Perceived Warmth and Competence. Perceived warmth refers to a person's perceived friendliness, likability, and sociability; perceived competence refers to a person's perceived intelligence, capability, and agency (Fiske, 2018). To assess users' perceptions of the chatbot's warmth and competence, we used the likability and perceived intelligence scales from the German version of the Godspeed Questionnaire (Bartneck et al., 2009; Cameron et al., 2021), comprising five items each. There are alternative scales that emphasize other facets of perceived warmth and competence, e.g., (B.-K. Lee et al., 2023) operationalized perceived warmth and competence using scales initially developed to measure subscales of trusting beliefs (McKnight et al., 2002) and Zheng et al. (2023) used a perceived warmth scale that also captures the perceived sociability during the interaction. However, warmth and likability as well as competence and intelligence, respectively, are conceptually very similar cognitive responses to new actors (Cameron et al., 2021; Fiske et al., 2007; Wojciszke et al., 2009). In addition, the items of the Godspeed questionnaire scales were explicitly developed to assess key concepts of human-robot interaction and are commonly used in HCI research (Weiss & Bartneck, 2015). Therefore, participants were asked to rate their impressions of the chatbot on seven-point semantic differential scales. Likability items included "unlikable–likable," "unfriendly–friendly," "impolite–polite," "agreeable–disagreeable," and "awful–nice". Perceived intelligence items included "incompetent–competent," "ignorant–educated," "irresponsible – responsible," "unintelligent–intelligent," and "unreasonable–reasonable".

Satisfaction. We assessed satisfaction with the chatbot using three items by Cheng and Jiang (2020). The items, i.e., "I am satisfied with the chatbot," "The chatbot did a good job," and "The chatbot did what I expected," were translated to German, adapted to chatbots, and measured on 7-point Likert-type rating scales (1 = *do not agree at all*, 7 = *completely agree*).

Intention to Use. Intention to continuously use the chatbot was measured using a validated scale by Cheng and Jiang (2020), comprising three items, i.e., "I would continue to use the chatbot," "I would use the chatbot for purposes other than my current use," and "I would explore other services of the chatbot than the ones I currently use". Continued use was translated to German, adapted to chatbots, and measured on a 7-point Likert-typed rating scale (1 = *do not agree at all*, 7 = *completely agree*).

Emotional Connection with the Company. Emotional connection with the company refers to "the affinity between customers and [a] brand" (Christodoulides et al., 2006, p. 805). We adapted Christodoulides et al. (2006)'s three-item scale by leaving out the item "I feel related to the type of people who are [X]'s customers" and replacing "brand" with our company

name in the other two items because we were interested in the emotional connection with the company (Araujo, 2018). The items “I feel like myclothes.deu actually cares about me” and “I feel that myclothes.deu really understands me” were translated to German and measured on 7-point Likert-typed rating scales (1 = *do not agree at all*, 7 = *completely agree*).

Original measurement items are provided in Appendix B. Descriptive statistics, bivariate correlations, and Cronbach’s alpha values for all study variables are displayed in Table 5.

Table 5 Descriptive Statistics, Correlations, and Alpha Values for Study Variables

Variable	<i>M</i>	<i>SD</i>	α	1	2	3	4	5
1. Perceived warmth	5.43	1.05	.93	—				
2. Perceived competence	4.87	1.05	.87	.67	—			
3. Satisfaction	5.03	1.44	.93	.61	.74	—		
4. Intention to use	4.20	1.60	.88	.55	.63	.74	—	
5. Emotional connection with the company	4.16	1.48	.92	.60	.70	.59	.64	—

Note. $N = 416$. Cronbach’s α . All correlations $p < .001$.

Manipulation Checks. We included manipulation checks for interaction mechanism and error response to ensure the manipulations were effective. To ensure participants encountered at least one chatbot error, we asked them how many mistakes they experienced during the interaction. To assess the error response manipulation, we asked participants whether the chatbot apologized for its error and disclosed that it was still learning at some point during the interaction.

3.4 Results

3.4.1 Manipulation Checks

The analysis of conversation protocols and self-reported data confirmed that all participants in the final sample experienced at least one error. Most participants indicated having encountered one or two errors (61.1%), while the others encountered three or more. Most participants in the social error response conditions noticed the chatbot’s social error response; 9 participants (2.2%) did not notice the chatbot’s apology, and 57 participants (13.7%) did not notice the chatbot’s disclosure that it was still learning.

Of all participants, the majority ($n = 278$) solved either four or five of five tasks, 87 solved three, 28 participants solved two, and 13 participants solved one task. The tasks that were solved most often were task 1 “gift card balance” ($n = 403$), and 3 “parcel tracking” ($n = 364$). Tasks 2 “delivery issue” and 4 “return cost” with the inbuilt error as well as the voluntary

task 5 “newsletter registration” were solved less often ($n_{\text{Task2}} = 334$, $n_{\text{Task4}} = 284$, $n_{\text{Task5}} = 222$). Participants mostly needed one attempt to solve tasks 1 (95.53%), 3 (83.79%), and 5 (96.40%), and two attempts for tasks 2 (61.68%) and 4 (72.89%).

3.4.2 Hypothesis Testing

To test H1–H3, we conducted five two-way analyses of variance (ANOVA) with interaction mechanism and error response as independent variables, including their interaction. We added warmth (a), competence (b), satisfaction (c), intention to use (d), and emotional connection with the company (e) as respective dependent variables (Table 6). Table 7 shows the means and standard deviations of the outcome variables by interaction mechanism and error response.

Table 6 Results of Two-Way Analyses of Variance in Study Variables

Effect	Perceived warmth		Perceived competence		Satisfaction		Intention to use		Emotional connection with the company	
	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}	<i>F</i> ratio	η^2_{partial}
Interaction mechanism	12.20***	.03	22.67***	.05	36.23***	.08	13.84***	.03	4.92°	.01
Error response	3.34°	.01	2.18	.01	4.89*	.01	2.65	.01	1.06	.003
Interaction mechanism × Error response	0.34	.001	0.21	.001	0.07	.0002	0.04	.0001	5.66*	.01

Note. $N = 416$. Interaction mechanism: 0 = button and 1 = free text. Error response: 0 = neutral and 1 = social. $df = 1,412$.

° $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 7 Means and Standard Deviations of Interaction Mechanism and Error Response Main Effects on Study Variables

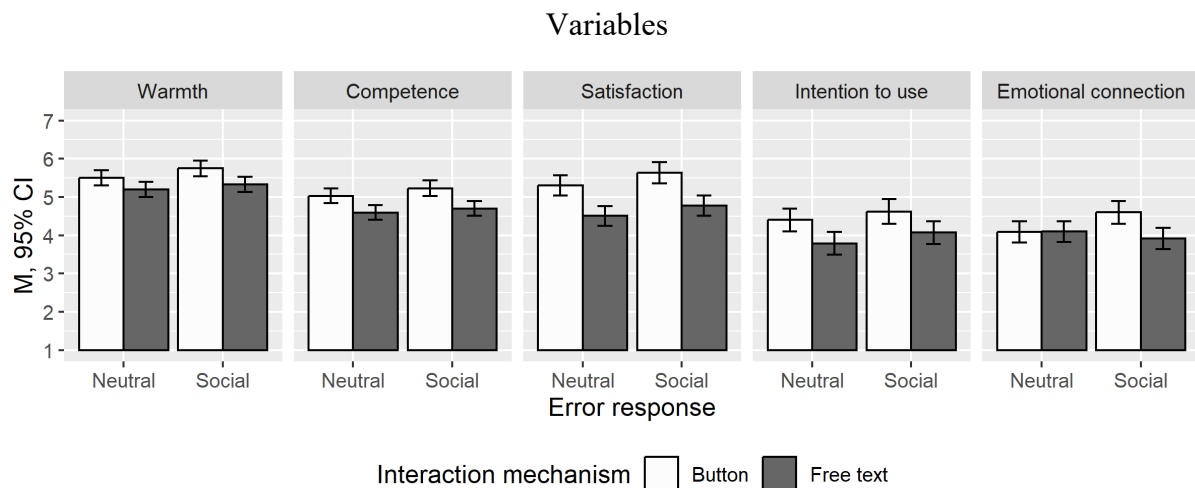
Variable	Interaction mechanism				Error response			
	Button		Free text		Neutral		Social	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Perceived warmth	5.62	0.98	5.26	1.08	5.35	1.06	5.52	1.03
Perceived competence	5.12	0.95	4.65	1.09	4.81	1.07	4.95	1.02
Satisfaction	5.46	1.21	4.64	1.53	4.90	1.52	5.18	1.34
Intention to use	4.50	1.56	3.93	1.60	4.09	1.63	4.33	1.57
Emotional connection with the company	4.33	1.43	4.01	1.52	4.09	1.40	4.23	1.57

Note. $N = 416$. Interaction mechanism: 0 = button and 1 = free text. Error response: 0 = neutral and 1 = social.

Contrary to our first hypothesis, we did not find positive main effects of interaction mechanism (free text) on perceived warmth and competence, satisfaction, and intention to use. Instead, participants indicated statistically significant lower values in the free text (vs. button) conditions on all dependent variables. Thus, H1 was not supported.

In line with our second hypothesis, we found a significant positive effect of social error response on satisfaction, $F(1, 412) = 4.89, p = .028, \eta^2_{\text{partial}} = .01$. Interacting with a socially responding chatbot led to significantly higher satisfaction ($M = 5.18, SD = 1.34$) than interacting with a neutrally responding chatbot ($M = 4.90, SD = 1.52$). The effect of social error response on perceived warmth was descriptively positive but only marginally significant ($M_{\text{Neutral}} = 5.35, SD_{\text{Neutral}} = 1.06, M_{\text{Social}} = 5.52, SD_{\text{Social}} = 1.03$), $F(1, 412) = 3.34, p = .068, \eta^2_{\text{partial}} = .01$. Although descriptively positive, error response had no significant effects on intention to use and emotional connection with the company. Only H2c was supported.

Regarding our interaction hypothesis (H3), we found one statistically significant interaction effect on emotional connection with the company, $F(1, 412) = 5.66, p = .018, \eta^2_{\text{partial}} = .01$, which was, however, not in line with our expectations (Figure 7).

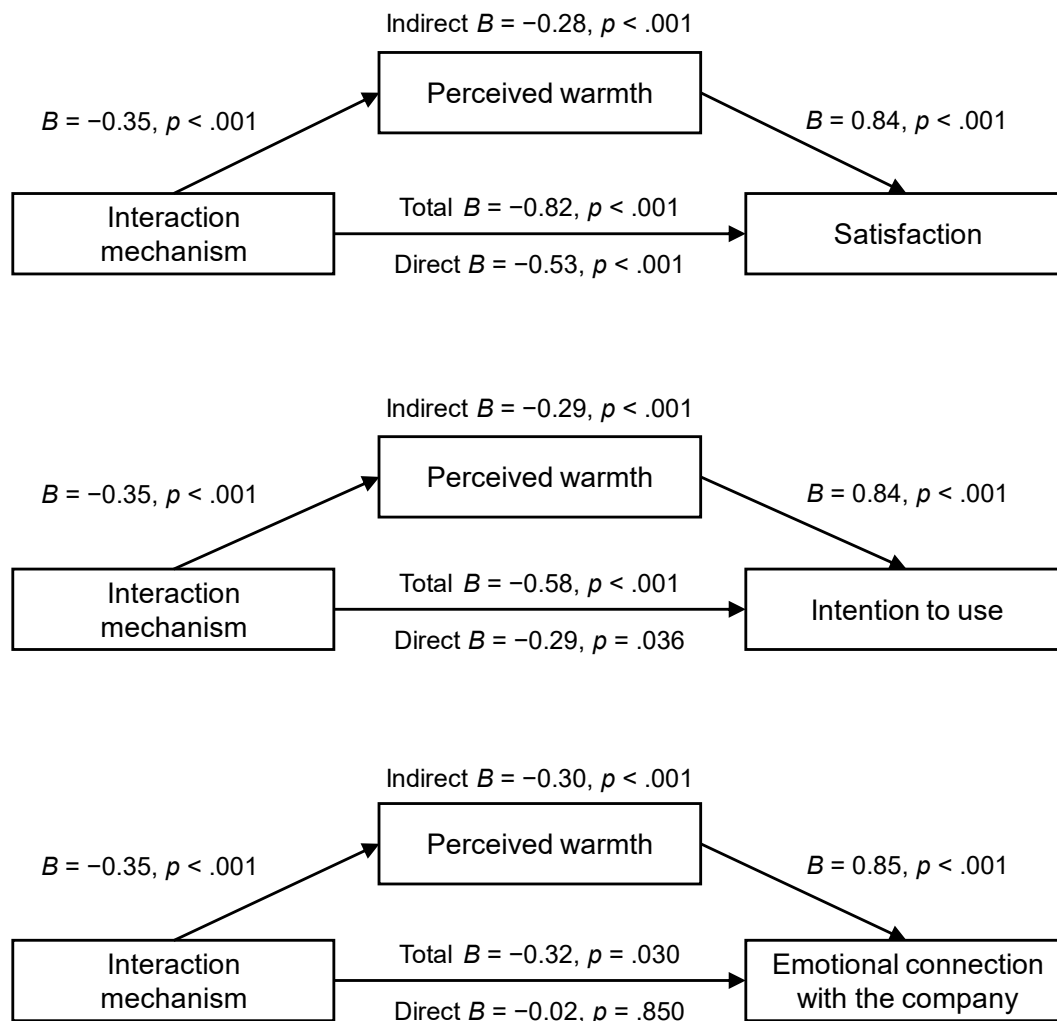
Figure 7 Interaction Plots of Error Response and Interaction Mechanism on Study

Note. $N = 416$.

We expected the effect of social (as opposed to neutral) error responses on the outcome variables to be stronger when the interaction mechanism is free text (as opposed to buttons). However, post hoc testing using Tukey's HSD only yielded one significant mean difference: Participants in the social error response conditions indicated a higher emotional connection with the company when interacting with the button-based chatbot ($M = 4.60$, $SD = 1.46$) compared to the free text chatbot ($M = 3.92$, $SD = 1.60$, $p = .006$, $d = -0.44$). No difference was found in the neutral conditions ($M_{\text{Button}} = 4.09$, $SD_{\text{Button}} = 1.36$, $M_{\text{Free text}} = 4.10$, $SD_{\text{Free text}} = 1.45$, $p = 1.000$). H3 was thus not supported.

To test H4 and H5, we ran separate mediation models for each outcome variable with interaction mechanism and error responses as independent variables, using the R *mediation* package v4.5.0 (Tingley et al., 2019). The mediation analyses conducted to test H4.1, which stated that perceived warmth mediated the effects of interaction mechanism on satisfaction, intention to use, and emotional connection with the company, yielded some interesting results (Figure 8). The interaction via free text was perceived as less warm, which led to less satisfaction, intention to use, and emotional connection with the company. Besides this indirect effect, there was still a direct effect indicating that warmth only partially explained the relationship. The same pattern emerged for the interaction mechanism effect on intention to use via perceived warmth, where a significant negative indirect effect emerged as well. Including the mediator in the model led to a negative direct effect that was smaller in magnitude than the total effect. Although there was no direct effect, we found an indirect effect from interaction mechanism on emotional connection with the company.

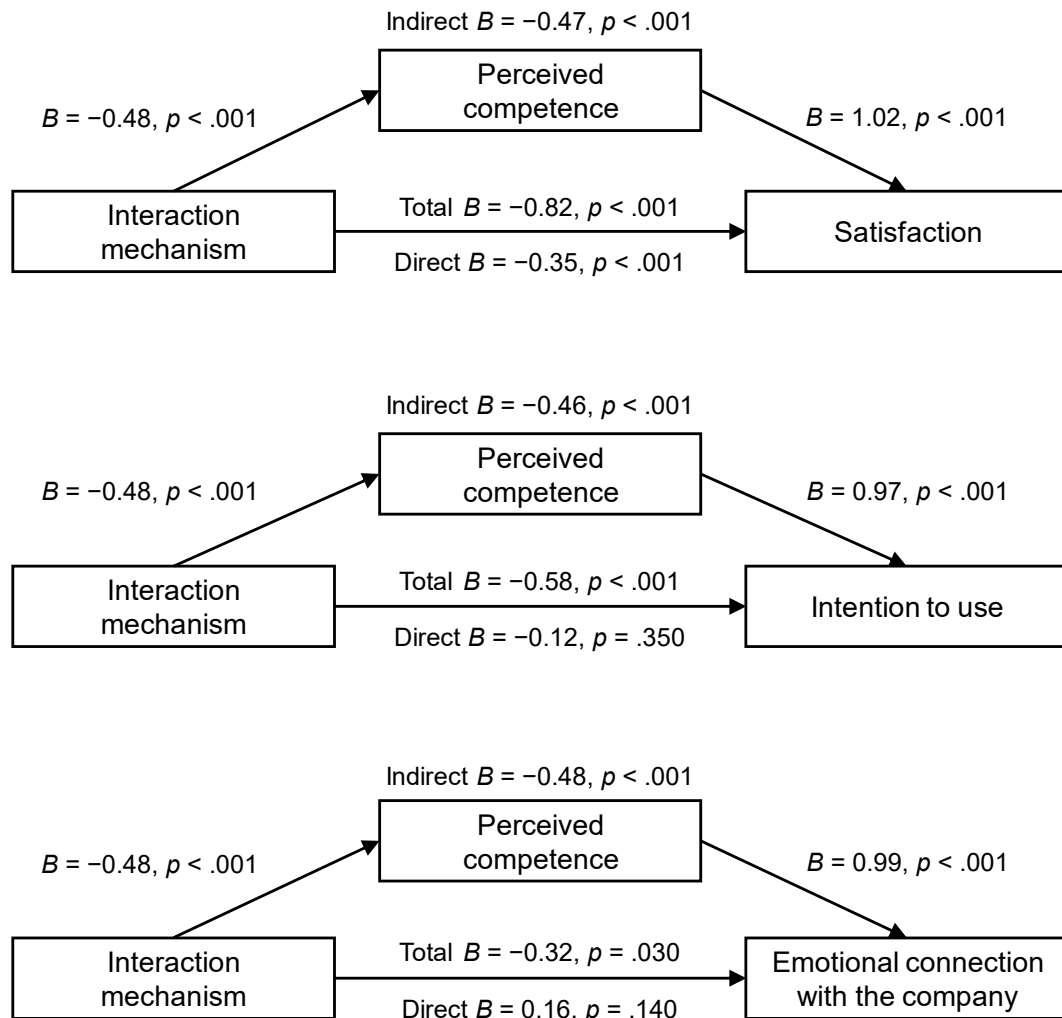
Figure 8 Mediation Results of Interaction Mechanism on Satisfaction, Intention to Use, and Emotional Connection with the Company via Perceived Warmth



Note. Interaction mechanism: 0 = button and 1 = free text. $N = 416$.

Figure 9 shows the results of the mediation analyses to test H4.2. Regarding perceived competence as mediator of the relationship between interaction mechanism and satisfaction, a significant negative indirect effect emerged, with the direct effect still being significant, indicating that perceived competence only partially explained the relationship between interaction mechanism and satisfaction. A full mediation was observed for intention to use and emotional connection with the company, where we found significant negative indirect effects as well as non-significant direct effects.

Figure 9 Mediation Results of Interaction Mechanism on Satisfaction, Intention to Use, and Emotional Connection with the Company via Perceived Competence



Note. Interaction mechanism: 0 = button and 1 = free text. $N = 416$.

A partial mediation with marginally significant indirect and total effects emerged when including perceived warmth as mediator of the relationship between error response and satisfaction (H5.1). We computed mediation models for the other outcomes, although no significant total effects were found when testing H2. No significant indirect effects were found for the other outcomes. Also, no significant indirect effects were found when including perceived competence as a mediator (H5.2). The full mediation models for H5.1 and H5.2 are displayed in Appendix B, Figures B.2 and B.3.

An overview of the results can be found in Table 8.

Table 8 Overview of Results

Hypothesis	Result
H1. Interaction mechanism → (a) perceived warmth and (b) competence of, (c) satisfaction with, (d) intention to use the chatbot, and (e) emotional connection with the company	Rejected; reverse effects
H2. Error response → (a) perceived warmth and (b) competence of, (c) satisfaction with, (d) intention to use the chatbot, and (e) emotional connection with the company	Partially confirmed (H2c)
H3. Interaction mechanism × error response → (a) perceived warmth and (b) competence of, (c) satisfaction with, (d) intention to use the chatbot, and (e) emotional connection with the company	Rejected
H4.1. Interaction mechanism → perceived warmth → (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company	Rejected; significant indirect effects, but reverse direction
H4.2. Interaction mechanism → perceived competence → (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company	Rejected; significant indirect effects, but reverse direction
H5.1. Error response → perceived warmth → (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company	Rejected
H5.2. Error response → perceived competence → (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company	Rejected

3.4.3 Additional Analyses

As the pretest results indicated, perceived ease of use might be a potential influencing factor of users' chatbot acceptance. Exploratory Welch two-sample t-tests yielded a higher perceived ease of use among participants who interacted via buttons compared to free text ($M_{\text{Button}} = 5.41$, $SD_{\text{Button}} = 1.20$, $M_{\text{Free text}} = 4.74$, $SD_{\text{Free text}} = 1.49$), $t(406.90) = 5.03$, $p < .001$, $d = 0.49$. Participants who interacted via buttons (vs. free text) also indicated fewer encountered errors ($M_{\text{Button}} = 1.89$, $SD_{\text{Button}} = 0.74$, $M_{\text{Free text}} = 3.31$, $SD_{\text{Free text}} = 2.03$), $t(276.58) = -9.65$, $p < .001$, $d = -0.92$. When controlling the relationship between interaction mechanism and the outcome variables for perceived ease of use, the negative effects of the free text interaction were mitigated: The effects on warmth, competence, intention to use, and emotional connection with the company became non-significant. The effect on satisfaction became less negative. Full results of the additional analyses can be found in Appendix B, Tables B.1-B.3.

3.5 Discussion

The rapid development of AI-based technologies has enabled chatbots to interact with users in a natural and human-like way. However, conversing with chatbots via free text often results in users experiencing feelings of frustration and abandoning conversations (Følstad & Taylor, 2020). To investigate the impact of interaction mechanism and error response, we conducted an online experiment where participants directly interacted with a chatbot in a common customer service setting. Four main findings emerged.

First, the free text chatbot was rated more negatively than the button chatbot regarding perceived warmth, competence, user satisfaction, and intention to use, which does not correspond with our expectations. This finding also contradicts previous research, which found free text interaction to positively impact service encounter satisfaction through feelings of humanness in the interaction (Diederich et al., 2019). Strikingly, additional analyses showed that the negative effects of interaction mechanism on the outcomes of interest were mitigated when controlled for perceived ease of use. This suggests that users are well trained through previous online experiences in interacting by mouse and hence preferred the button chatbot, which was navigable using the mouse. This is directly related to the fact that typing a request typically takes longer than simply clicking a button. Users might have preferred the button chatbot because the buttons matched the tasks very well, were easy to find, and structured the problem, whereas free text input requires the user to come up with the proper words, questions, and requests, as our exploratory analyses suggest. When users entered their request in this study, they could either do so via pre-defined buttons or via free text, i.e., only one interaction mechanism was available at the same time. Future research could investigate whether allowing users to choose between buttons and free text input would be beneficial, as this provides guidance and flexibility simultaneously (Haugeland et al., 2022).

Second, social error responses increased users' satisfaction with the chatbot (H2). Our study conceptually replicated previous findings on the positive impact of social error responses on user satisfaction with social robots in the domain of text-based chatbots (Cameron et al., 2021). However, we did not find significant effects of social error responses on perceived warmth, competence, intention to use, and emotional connection with the company.

Surprisingly, we did not find interaction effects in line with our third hypothesis. There was, however, one significant interaction effect: The free text chatbot that responded socially to errors lead to a lower emotional connection with the company than the button chatbot that responded socially to mistakes, whereas the two chatbots did not differ when not responding socially to errors. A reason for that could be that the button (vs. free text) chatbot, which

apologized for its errors and, at the same time, communicated that it is still learning, might have signaled to users that the company tries to anticipate its customers' needs by collecting frequently asked questions and turning them into buttons. However, since we did not preregister this effect, its interpretation is exploratory and must be taken cautiously. Overall, we do not find evidence that the effects of social error responses on warmth, competence, satisfaction, and intention to use depend on how users interact with a chatbot.

Fourth, perceived warmth and perceived competence mediated the effect of interaction mechanism on the outcomes. However, mediation effects did not go in the predicted direction (H4). Instead, as suggested by our results for H1, free text interaction predicted less satisfaction, intention to use, and emotional connection with the company, presumably because participants perceived the free text chatbot to be less warm and less competent. The result of our additional analyses, namely, that perceived ease of use mitigated the negative effects of free text interaction as well as the fact that perceived ease of use was strongly correlated to both perceived warmth ($r = .54$) and competence ($r = .59$), suggest that the lower the perceived ease of use of a chatbot, the less warm and competent users perceive the chatbot. We did not find perceived warmth and competence to be significant mediators of the relationship between error response and the outcomes.

3.5.1 Theoretical Implications

The MASA paradigm and Social Response Theory claim that the more social or human-like a chatbot is, the more positive the user responses will be (Lombard & Xu, 2021). However, this prediction appears to discount the relevance of usability aspects to the extent that they matter more than previously thought: Our results suggest that the benefits of using buttons, i.e., structuring the conversational flow and enhancing navigability, appear to outweigh the benefits of human-like interactions with chatbots. From a theoretical point of view, our findings thus challenge the applicability of the MASA paradigm in task-based contexts like customer service. There might be a difference between the effects of a human-like interaction mechanism and other human-like cues of the chatbot, e.g., language style or personality. Our study suggests that while human-like language cues generally positively impact user responses (Adam et al., 2021; Araujo, 2018), a human-like interaction mechanism might have different effects because it might be perceived as inefficient and impractical in particular in reasonably structured problems such as customer service (Haugeland et al., 2022).

The findings may also be an indication that the preference for button vs. free-text bots is contingent on the type of tasks the user has to solve. The tasks we studied could have been more suitable to be answered based on buttons (vs. free text). It would be interesting to

investigate which questions users believe the button and free text bots are suitable for, i.e., buttons might be more suitable for task-related queries, and free text might be more suitable for emotional/social topics.

Although participants in this study interacted with an actual chatbot, the scenario was fictitious, i.e., the conversation was not about an actual order, reducing participant stakes. However, solving real-world tasks in the context of customer service, e.g., keeping track of a parcel, may also be associated with emotional responses, such as anger or frustration. Future research could, therefore, explore the impact of the interaction mechanism in contexts where there is more at stake for the user, i.e., where they need to explain, argue, or justify more complex matters and situations or in situations where an active exchange and social factors matter more, e.g., in mental health or financial counseling.

Our results on the effect of social error responses on satisfaction align with the main premise of social response theory, i.e., the more social cues a media agent possesses, the more positive users respond to it (Nass & Moon, 2000). We did not find significant effects of social error responses on the remaining outcomes. Language cues are secondary cues, i.e., “they have less power in activating user’s social perception and responses” (Lombard & Xu, 2021, p. 34). The apology, combined with the statement that the chatbot is still learning, may not have been salient enough to elicit feelings of similarity with the chatbot. This could be why we did not find the expected effects for perceived warmth, competence, intent to use, and emotional engagement with the company. At the same time, the wording we chose reflects what is commonly used in real chatbot applications. Hence, it appears that the social error response does not seem to be the decisive factor that governs user perceptions of chatbots. Our mediation findings indicate that perceived warmth and competence are indeed central constructs for people’s evaluations of a chatbot (Fiske et al., 2007).

3.5.2 Practical Implications and Outlook

Our findings have several implications for marketing practice (e.g., M.-H. Huang & Rust, 2021b). A direct corollary of the results is that firms seeking to implement chatbots may want to leverage the benefits of both interaction mechanisms, i.e., buttons can structure the conversational flow of a chatbot when the problem is well-defined, and free text input might help with new user queries or queries that occur less frequently. Hence, the findings suggest that companies should give relatively more weight to the usability aspect over the potential benefit of human-like interaction mechanisms. In contrast, firms need to be less concerned about the way they implement an error response, i.e., consumers appear to appreciate a social

error response, but we do not have evidence that this is a strong driver of how they view the chatbot or the firm.

It is also of relevance for firms to understand how these findings relate to the role of the recently released LLMs like GPT-4 (OpenAI, 2023). LLMs have the potential to provide even more efficient and accurate automated customer service, e.g., by providing personalized, human-like responses to customer requests (Floridi & Chiriatti, 2020; Rivas & Zhao, 2023). In this light, we must ask ourselves if our results will look different in a few years when people have more experience using chatbots powered by LLMs. However, several reasons speak for the continued relevance of our results:

LLMs have to be trained on current and relevant customer data: Until timely data is available to train these models, they are not a real alternative for companies to current customer service chatbots (Rivas & Zhao, 2023). In our experiment, participants had to solve pre-defined tasks. However, knowledge of chatbots based on LLMs is often rather broad than deep, i.e., they lack knowledge on specific topics. ChatGPT, for example, is often used for writing poems in different styles (Haque et al., 2022) but not necessarily for specific tasks like booking a flight or registering for a degree program. Yet, the primary goal in customer service is to give customers controllable, complete, and correct information (Rivas & Zhao, 2023).

ChatGPT's interaction mechanism is free text. However, developers of LLM-based chatbots also employ buttons; e.g., Bing's chatbot suggests answer options and further questions via buttons (Bing, 2023). Therefore, researching the influence of different interaction mechanisms remains a relevant line of research. Even chatbots based on LLMs make mistakes or hallucinate, e.g., they give incorrect answers to math problems, make up references, or get tangled up in contradictions (Borji, 2023; Sundar & Liao, 2023; Van der Meulen, 2023). Thus, the need for an adequate response remains relevant, e.g., when a user confronts the system with an error it has made or is programmed not to answer specific questions.

Lastly, although LLMs provide many opportunities, they come with less control over the conversational flow. Our collaboration with a startup specialized in conversational AI gave us full control over our experimental stimuli, i.e., the content, timing, and type of chatbot utterances, a prerequisite for valid and reliable experimental research on humanmachine communication (Greussing et al., 2022). At the same time, this high degree of control over the answers is also critical for firms that must ensure that the chatbot's responses are always accurate (Rivas & Zhao, 2023). Hence, the lack of control over the conversational flow is not only a challenge for valid and reliable experimental studies but also for firms seeking to

implement this kind of technology. This challenge is amplified by recent findings of substantial variability of LLM's responses over time (L. Chen et al., 2024).

Chatbot quality is a potential moderator of the relationships between its social cues and users' responses. Although the transcripts of our participants' chat conversations indicate that this chatbot was already quite good, we see the need for future research comparing state-of-the-art chatbots with chatbots based on LLMs.

3.5.3 Limitations and Future Research

First, experimental online studies have higher internal but less external validity than field studies. We, therefore, suggest future research to validate our findings in a real-world environment. Compared to our study's controlled experimental setting that included a fictitious scenario and fake customer credentials, customer stakes and potentially also emotions such as anger and frustration, are higher when interacting with a customer service chatbot about a real order, e.g., when a package has been lost on the way to the customer. Emotions like customer anger might moderate the effect between the predictors and user acceptance. Specifically, social cues might lose their benefits when customers already enter the interaction in an angry state (Crollic et al., 2022). In this context, qualitatively analyzing chat transcripts of interactions between customers and chatbots might provide valuable additional insights into how users talk to chatbots, how they immediately react to a chatbot apologizing for an error, and when and why users abort a conversation. Second, we investigated our research question in the customer service context, as chatbots are already commonly used here. However, there might be differences in the impact of free text interaction or social error responses regarding chatbots in domains where social and emotional factors matter more, e.g., chatbots as social companions or mental health advisors. In addition, chatbots in customer service frequently answer relatively simple and predictable questions customers might have before or after purchasing or using products or services. Future research could, therefore, address contexts in which users face more complex problems. Third, all our study participants were based in Germany. Testing our hypotheses using a sample from a different cultural context would further strengthen our results. Finally, we cannot make any statement about causality regarding the mediation analyses. A longitudinal study could provide answers to this question in future research.

3.6 Conclusion

This study investigated the impact of free text interaction and social error responses in human-chatbot interactions on perceived warmth and competence, satisfaction, intention to use, and emotional connection with the company through an online experiment. Interestingly, we found that chatbots using free text interaction decrease users' warmth and competence

perceptions, satisfaction, intention to use, and emotional connection with the company compared to chatbots with a structured conversation flow. Additionally, chatbots that respond to errors in a social way increase user satisfaction but not the other outcomes. Our findings suggest that designing chatbots to be more human-like is not always optimal for user perceptions of the chatbot and the employing company, particularly when the chatbot is strongly task-focused. In our study, the chatbot's error-proneness trumped its humanness. Our insights can guide practitioners in designing efficient, accurate, and accessible chatbots.

4. A Meta-Analysis of the Effects of Social Cues on Social Responses

Statement of Contribution

Chapter 4 presents a slightly modified version of the single-authored manuscript, titled „The Effects of Human-Like Social Cues on Social Responses Towards Text-Based Conversational Agents—A Meta-Analysis”. I, Stefanie Helene Klein, am the sole author of this chapter. I confirm sole responsibility for research questions and design, data collection, preparation, analysis and interpretation of the results, as well as for manuscript preparation.

An extended abstract has been presented at the 74th Annual Conference of the International Communication Association (ICA). The full manuscript is under review at a peer-reviewed journal.

Acknowledgements

I thank Anne Bucher for her assistance in coding the records.

Abstract

This meta-analysis examines the effect of text-based chatbots' social cues on users' social responses and the influence of potential moderators. Experimental studies that manipulate human-likeness using social cues, examine user responses, including attitude, perception, affect, rapport, trust, and behavior, and compare human-like chatbots with chatbots having no or lower-degree human-like characteristics, were included. A systematic search for published and unpublished research resulted in a final sample of 800 effect sizes from 199 datasets reported in 142 papers ($N = 41,642$). Meta-analytic random-effects models computed overall, and for each outcome category yielded a small effect of human-likeness on social responses ($g = 0.36$, 95% CI [0.27, 0.44]). The results further suggested that human-like chatbot characteristics improve user responses to varying degrees and under different boundary conditions. The findings can guide practitioners in designing effective and ethically justifiable chatbots.

Keywords: meta-analysis, chatbot, conversational agent, social cue, social response, human-likeness

4.1 Introduction

Recent advances in technologies based on artificial intelligence (AI) have enabled an enormous surge in the use of text-based chatbots across various sectors, including service, sales, counseling, and consulting (Jacobson & Gorea, 2023; W. Liu & Yao, 2023). These chatbots are text-based software agents that communicate with users through natural language (Shawar & Atwell, 2007). Even before the public release of chatbots powered by large language models (LLM), developers have been striving to design chatbots that mimic human characteristics and behaviors (Silva & Canedo, 2024). They have equipped these chatbots with social cues, such as a smiling avatar, a casual conversation style, or delayed response times, with the goal of enhancing user responses, including agent evaluations and user engagement (Chaves & Gerosa, 2021; Feine et al., 2019; Nass & Moon, 2000).

Over the last ten years, there has been extensive empirical research on the influence of human-like characteristics on users' acceptance of intelligent agents (for a recent systematic review, see Rapp et al., 2021). Roesler et al. (2021) conducted a meta-analysis on the effectiveness of embodied robots' anthropomorphic design features on human-related outcomes and arrived at a moderate positive average effect. In their meta-analysis on the antecedents and consequences of users' perceived anthropomorphism in service provision, Blut et al. (2021) found that perceived anthropomorphism and intention to use were strongly correlated but text-based conversational agents were the research focus of merely eight percent of the included studies. However, compared to embodied conversational agents like the social robots Pepper or Cosmo, voice-based agents like Amazon's Alexa, and 3D avatars, text-based chatbots communicate exclusively over a "messaging-based interface" (Araujo, 2018, p. 184). Text-based chatbots are disembodied conversational agents because they usually have no or less physical or virtual embodiment, represented, for instance, by facial expressions, body movements, and gaze (Krämer et al., 2009). They do not rely on voice-based input and output, which does not allow the chatbot or the user to use expressions at different volumes and pitches, so communication is limited to exchanging text messages (Zierau et al., 2023). However, prior research suggests that text-based chatbots might still elicit humanness perceptions, even without using sensorial cues like voice and movement (Rapp et al., 2021). Thus, human-likeness effects might work differently in chatbots compared to more embodied agents.

Previous research generally suggests beneficial effects of human-likeness in text-based chatbots (e.g., Adam et al., 2021; Araujo, 2018; S. Lee et al., 2020). However, under certain circumstances, human-like characteristics can also harm user responses, for example, due to the chatbot triggering uncanny emotions (Ciechanowski et al., 2019), when users enter an

interaction in anger (Crolic et al., 2022), when the service interaction is flawed (Brendel et al., 2023) or when chatbot users are experienced (Gnewuch et al., 2022). Many survey and experimental studies exist regarding the effects of social cues in chatbots, leading to several review articles (e.g., the literature reviews by Bavaresco et al., 2020; Chaves & Gerosa, 2021; Van Pinxteren et al., 2020). With regard to text-based conversational agents, meta-analyses exist for lifestyle and mental health outcomes (e.g., Abd-Alrazaq et al., 2020; Singh et al., 2023), with the exception of Gopinath and Kasilingam (2023), who investigated the factors that influence intention to use service chatbots.

Due to the ambiguous effects of social cues on social responses, the character of communication between humans and disembodied chatbots, and because research on chatbots has greatly evolved in recent years, I see the need for a meta-analysis on the effects of social cues on users' responses towards text-based chatbots.

4.2 Main Effects of Human-Likeness

The independent variable of this study is a chatbot's human-likeness as expressed through the social cues that a chatbot is equipped with. Social cues can be visual (avatar-related, e.g., smile or chat-related, e.g., emoji use), verbal (related to content, e.g., small talk, or style, e.g., language use, speech style, message contingency), invisible (e.g., response times) (Feine et al., 2019), or related to the interaction (Zierau et al., 2020). The dependent variables of this meta-analysis are social responses, that is, outcomes that represent users' attitudes, perceptions, positive and negative affective, relational, trust-related, and behavioral reactions towards a chatbot, a chatbot message, or an interaction with a chatbot (G. Huang & Wang, 2023; Xu, Chen, et al., 2023). These social response categories emerged from an extensive literature review and discussions with academic experts in conversational agents and meta-analysis.

Attitude toward the chatbot or the interaction refers to users' general positive or negative evaluations of a chatbot, a message, or an interaction (Blut et al., 2021), encompassing concepts like chatbot acceptance, and evaluations regarding usability and user satisfaction. The impact of human-like cues on attitudes has been investigated in various studies (e.g., De Sá Siqueira et al., 2023; Gelbrich et al., 2021; S. Lee & Choi, 2017). Chatbot perception refers to how users think about or what impression they have of a chatbot, a message, or an interaction (G. Huang & Wang, 2023). Perceptions are often assessed using measures of anthropomorphism, humanness, intelligence, competence, expertise, safety, credibility, informativeness, and morality. Social cues have been found to influence chatbot perceptions in several studies (e.g., Go & Sundar, 2019; Schuetzler et al., 2020; Westerman et al., 2019). Positive affect represents the users' positive feelings directed at or triggered by a chatbot (Fishbach & Labroo, 2007) and

can be operationalized by perceived enjoyment, pleasure, warmth, arousal, valence, or social attraction (e.g., S. Lee & Choi, 2017; Meng et al., 2023; Westerman et al., 2019). Negative affect, in contrast, refers to users' negative feelings, for instance, feelings of discomfort, strain, eeriness, threat, or anger (e.g., Song & Shin, 2024; Zhang et al., 2022). Rapport relates to the personal relationship between user and chatbot (Wirtz et al., 2018) and is often assessed via measures of social presence, closeness, perceived intimacy, attachment, companionship, empathy, and parasocial relationship (e.g., Go & Sundar, 2019; Konya-Baumbach et al., 2023; S. Lee et al., 2020). Trust is defined as “the attitude that a chatbot will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability” (J. D. Lee & See, 2004, p. 51). Although trust has been treated as a subcategory of attitude in the past (Roesler et al., 2021), I treat it as a separate category due to the large expected amount of research. Trust is often measured in terms of trustworthiness or trust toward the chatbot (e.g., Park et al., 2024; Song & Shin, 2024; Wald et al., 2021). User behavior constitutes a separate category. Examples of measures are self-disclosure, task performance, prosocial behavior, behavioral intentions, and general adoption (e.g., Kang & Kang, 2024; Scherer et al., 2020).

According to the Media Are Social Actors (MASA) paradigm and social response theory, humans apply the same rules in their interactions with technological systems as in their interactions with other humans (Lombard & Xu, 2021; Nass & Moon, 2000). Technologies provide verbal and non-verbal social cues that can elicit social responses (Feine et al., 2019; Go & Sundar, 2019). The more social cues—“characteristics that are associated with humans” (Nass & Moon, 2000, p. 97)—a technology presents, the more likely humans are to act socially towards it. Prior studies generally indicated positive effects of human-like cues in chatbots (Araujo, 2018; Go & Sundar, 2019; S. Lee et al., 2020). However, human-like chatbot characteristics can also hurt social responses under certain circumstances. For instance, Ciechanowski et al. (2019) showed in their experiment on the impact of human-likeness on affective responses that a more human-like chatbot triggered more uncanny emotions, negative affect, and more intense psychophysiological reactions compared to a simple text-based chatbot. A series of experiments by Crollic et al. (2022) yielded negative human-likeness effects on customer satisfaction and purchase intentions when the customer entered the chatbot interaction angrily. Related, Brendel et al. (2023) concluded that human-like cues lead to user frustration in flawed service interactions. Previous chatbot experience might also hinder the effectiveness of social cues. For instance, Gnewuch et al. (2022) found that dynamically delayed (vs. instant) response times only improved user responses for novice chatbot users. Therefore, I aim to answer the following research question:

RQ1. What is the overall effect of text-based chatbots' social cues on users' social responses?

4.3 Moderators of Human-Likeness Effects

Based on the results of similar meta-analyses on human-likeness effects in robots, I expect high degrees of heterogeneity. The focus is on factors relating to the user, the agent, the context, and the method to investigate the conditions under which human-like characteristics potentially facilitate (or hinder) human-chatbot interaction. The moderators were identified through a comprehensive review of empirical research (see, for example, the meta-analyses by Roesler et al. (2021) and Xu et al. (2023) for robots, and Blut et al. (2021) for automated agents in service provision). I ask the following research question:

RQ2. What factors moderate the effects of text-based chatbots' social cues on users' social responses?

4.3.1 Sample Characteristics

The current research focuses on mean age, gender, and chatbot experience as potential sample-related moderators. Related to Scopelliti et al. (2005)'s finding that older adults consider new technologies more complicated to use, Chattaraman et al. (2019) concluded that older participants preferred a concise (vs. social) conversational style as well as an abstract (vs. humanoid) and static agent appearance to reduce distraction. It is thus possible that the human-likeness effects on social responses are smaller for older people as human-like features distract them from processing content-related information. Smaller effects for older adults are also plausible due to their lower tendency to anthropomorphize (Letheren et al., 2016). However, Straßmann et al. (2018) showed that human-like (vs. non-human-like) virtual agents received the highest use intention and bonding scores among seniors, while, among students, human-like agents received the lowest scores on both variables.

Regarding gender, people who identify as female are likely to be more skeptical of new technologies than male-identifying people (Scopelliti et al., 2005), which is why the effect of human-likeness on social responses might be generally larger for male users. On the contrary, prior research suggests that female-identifying persons appreciate a robot's human-like interaction and behavior more than male-identifying persons (Pelau et al., 2021), making plausible larger human-likeness effects for female-identifying persons.

Highly experienced chatbot users are likely very familiar with the technology and are used to being exposed to social cues (Gnewuch et al., 2022). As they might know more about the underlying technologies, the human-likeness effects might be smaller for this group.

4.3.2 Agent Characteristics

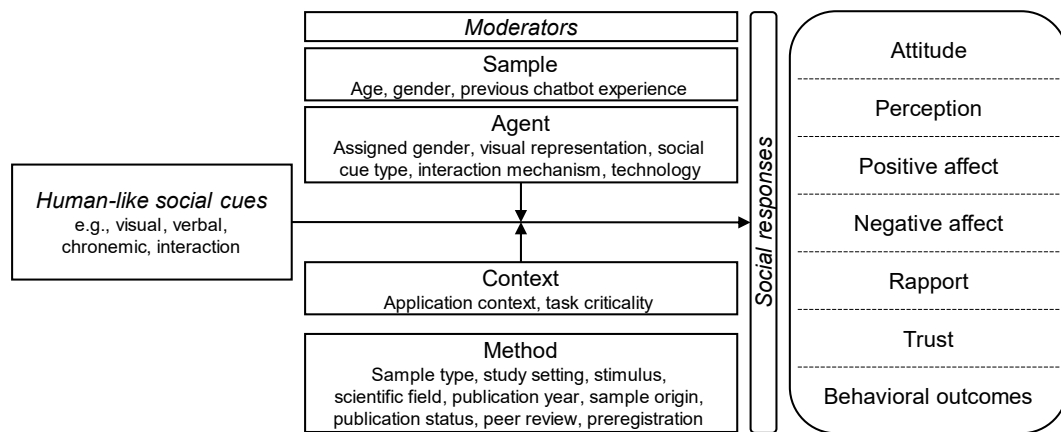
Whether agent characteristics moderate human-likeness effects on social responses is another question, this meta-analysis aims to answer. Previous meta-analyses included the agent's assigned gender as a potential moderator (Blut et al., 2021). According to Feine et al. (2020), most text-based chatbots possess gender-specific cues (e.g., names and avatars), of whom most imply female gender. Gender-specific cues may lead users to apply human gender stereotypes to chatbots. For instance, survey results by Kim Soomin et al. (2021) indicate that users expect different kinds of tasks from female and male chatbots, e.g., having a chat, advising, healthcare, and clerical work was expected to be carried out by female chatbots, while male chatbots were rather expected to do professional work. Human-likeness might be more effective for chatbots with female cues because they elicit stereotypes relating to likability and helpfulness (Koenig, 2018), enhancing the effect of human-like cues. Besides agent gender, I focus on agent characteristics specific to text-based chatbots, e.g., their visual representation, implementation of human-likeness, interaction mechanism, and technological foundation. The chatbot's visual representation (i.e., whether it has a human-like avatar) might affect the human-likeness effects on social responses. Previous research has shown that visual social cues matter (Go & Sundar, 2019), which aligns with MASA's proposition that cues that are more strongly connected to people's perceptions of an agent's socialness (i.e., primary cues like a human-like avatar) are more effective in eliciting social responses (Lombard & Xu, 2021). Therefore, I expect that social cues are more effective when chatbots have human-like avatars. As interactions with chatbots mainly rely on textual information, using verbal (vs. other) cues to manipulate the independent variable might increase human-likeness effects. Consequently, verbal cues might play a larger role than other cues, including visual, invisible, and interaction cues. In contrast, the differentiation between primary and secondary social cues might also come into play here, suggesting that secondary cues, such as language use, are not as efficient as other cue types (Lombard & Xu, 2021). The chatbot's interaction mechanism might also be a moderator. Structured conversational flows using predefined answer buttons, may appear less human-like than unstructured flows where interaction happens via free text (Diederich et al., 2019). Unstructured interactions may magnify the positive human-likeness effects and thus lead to improved social responses. Last, I look at the chatbot's technological foundation. Compared to simple rule-based chatbots, more elaborate technological bases, e.g., LLMs, enable more complex and thus more human-like conversations (Abercrombie et al., 2023), which might magnify the positive human-likeness effects.

4.3.3 Context Characteristics

Application context and task criticality might also moderate human-likeness effects. Effects might be larger in hedonic contexts as chatbot interactions here should provide pleasure and entertainment, focusing more on social aspects like relationships and emotions, compared to functional, more task-based contexts like customer service (G. Huang & Wang, 2023; Roesler et al., 2021). The criticality of tasks performed in a specific context refers to whether the stakes for users are high (Blut et al., 2021). Stakes for users can be high in insurance, banking, and mental health counseling, whereas stakes are generally lower for tasks like recipe or movie selection and chit-chat. Larger effects for critical contexts are plausible, as human-likeness might persuade people to use a chatbot more readily. However, human-likeness effects might be more effective in uncritical tasks due to users being more relaxed, and human-likeness contributing to a more natural interaction and a positive user experience. In addition, users might expect chatbots performing critical tasks to be more neutral to avoid distractions.

4.3.4 Method Characteristics

The current research also includes method characteristics that have been shown to influence meta-analytic results and are specific to research on media agents as potential moderators. Compared to non-student samples, student samples are more homogeneous in their composition, which can lead to larger effect sizes (Orsingher et al., 2010). Laboratory experiments provide larger control over participant behavior than online and field experiments, and exposing participants to an actual interaction experience with a chatbot might be considered more realistic than showing participants prerecorded screenshots or videos, potentially resulting in larger effect sizes (Greussing et al., 2022). The year of publication (as chatbot experience might increase with time), the sample country (i.e., whether the research was conducted in the US) (Blut et al., 2021), and the scientific field (i.e., whether the paper was published in a communication science outlet) were also considered as potential moderators. I assessed potential publication bias by considering publication status and whether a paper has been peer-reviewed. Further, I included whether a study was preregistered, as non-preregistered studies could have larger effect sizes than preregistered ones (Van den Akker et al., 2023). Figure 10 shows the analytical framework for this meta-analysis.

Figure 10 Analytical Framework

4.4 Method

This meta-analysis followed the preferred reporting items for systematic review and meta-analysis (PRISMA) framework (Page et al., 2021), with a preregistered PRISMA protocol (Moher et al., 2015; Shamseer et al., 2015) on the Open Science Framework (OSF): https://osf.io/9t2q8?view_only=b87fe963d36e4ed586865fdfe880a3e3. I adapted the templates provided by Moreau and Gamble (2022) for reporting search queries, search results, extracted data, and deviations from the protocol, as well as for the call for unpublished work. The deviations from the protocol can be found in Appendix C, Table C.1.

Supporting data and corresponding R code files are accessible at https://osf.io/cezmu/?view_only=48d1e1ccb7b6419294ea54f6fb24cd66. Since a paper (e.g., a journal article or conference paper), may contain the results of several experiments, I use the term dataset to distinguish between different data collections within a paper throughout the method and results sections (Van Berlo et al., 2023).

4.4.1 Literature Search and Selection

The eligibility criteria were developed following the population, intervention, comparison, and outcomes (PICO) framework (Richardson et al., 1995). Population: I included samples of healthy adults, where healthy refers to the absence of disease or impairment. Intervention: Studies manipulating human-likeness using one or multiple social cues were eligible. Social cues included visual (e.g., smile, emoji use), verbal (e.g., small talk, speech styles), or invisible (e.g., response time) social cues (Feine et al., 2019; Lombard & Xu, 2021). Comparison: I included studies comparing human-like text-based chatbots with those with no or lower-degree human-like, technical, or machine-like characteristics. Outcome: Studies that assess one or multiple dependent variables relating to attitude, perception, positive and negative affect, rapport, trust, and behavioral outcomes were eligible.

A systematic literature search in the electronic databases Web of Science, ERIC, PsycInfo, Business Source Premier, EconLit, Academic Search Premier, IEEE Xplore Digital Library, ACM Digital Library, Open Access Theses and Dissertations (OATD), OpenDissertations, Networked Digital Library of Theses and Dissertations (NDLTD), and Google Scholar (first 500 search results) was carried out. All records to which I had access by September 15, 2023, were considered. The terms used for the literature search included combinations of chatbot, human-likeness, empirical study, and synonyms thereof. A specific query for each database was developed. Other search activities included scanning reference lists of included papers or prior reviews (i.e., Abercrombie et al., 2023; Chaves & Gerosa, 2021; Rapp et al., 2021; Van Pinxteren et al., 2020). Additionally, I sent requests for unpublished research to discipline-specific mailing lists and the author's professional network. Appendix C contains the literature search terms (Table C.2), and an example search query. The literature search and screening protocol as well as the search queries for each database can be accessed at https://osf.io/cezmu/?view_only=48d1e1ccb7b6419294ea54f6fb24cd66.

After removing duplicates, the title, abstract, and keywords of the search results were screened manually using the free online tool Abstrackr (B. C. Wallace et al., 2012). Initially, two coders screened a sample containing the same 100 records (inclusion and exclusion criteria in Appendix C, Table C.3). Cohen's Kappa for the categories 'excluded' vs. 'not excluded' was computed. "Not excluded" records included records tagged as "included" and "uncertain" that would later be eligible for a full screening. We solved disagreements through discussion. A moderate Cohen's Kappa of .463 resulted, which I considered sufficient. Thus, two coders rated all identified records. We included records 1) if they relied on empirical research, 2) on text-based chatbots, 3) with human-likeness as an independent variable, 4) and at least one social response as the dependent variable. We excluded records when retracted, not in English or German, or when they referred to embodied robots, augmented reality avatars, or voice-based agents. Studies on children, conference prefaces, newspaper or blog articles, reviews, and qualitative papers were also excluded. I then conducted the second screening of the search results, considering the full eligibility criteria: Records had to report survey or experimental research on samples of healthy adults, where healthy referred to "the absence of disease or impairment" (Sartorius, 2006, p. 662). Records had to manipulate human-likeness using one or multiple social cues and compare human-like text-based chatbots with chatbots having no or lower-degree human-like, technical, or machine-like characteristics. Records needed to assess single-scale items or indices representing attitude, perception, positive and negative affect, rapport, trust, or behavioral outcomes. I included published manuscripts, contributions in

conference proceedings, dissertations, final theses, preprints, and unpublished manuscripts in English and German. Last, records had to report sufficient statistical results to allow the calculation of effect sizes. If possible, I obtained full papers for all records that fulfilled all inclusion criteria. The flow of papers through the screening process is shown in Appendix C, Figure C.1.

After screening 4,719 records, 510 full texts were assessed for eligibility, and 142 papers from 2007 to 2023 formed the final sample (Appendix C). All final papers reported experimental research. The majority were journal articles (59.9%), 27.5% were conference papers, 8.5% were theses, and 4.2% were preprints and unpublished manuscripts⁹. Papers came from the fields of human-computer interaction (26.1%), business (19%), information systems (14.8%), psychology (13.4%), communication (9.9%), and computer science (7%). About 10% of papers were located in an interdisciplinary context. Most papers were published at the time of submission (97.9%), of whom 89.1% had been peer-reviewed.

Information for the computation of effect sizes and variances, e.g., Cohen's d , means, and standard deviations, and regarding the moderators were systematically extracted and coded by the author and a student assistant in Microsoft Excel. The moderator variables extracted and used in this meta-analysis are shown in Table 9.

4.4.2 Effect Size Calculation

First, I calculated the standardized mean difference Cohen's d , and its variance for every effect identified in the extraction process (Borenstein & Hedges, 2019; Lenhard & Lenhard, 2022; Wilson, 2023). As Cohen's d has been shown to have a slight positive bias in small samples, Hedges' g , its corrected, unbiased estimate, was used (Borenstein & Hedges, 2019). Hedges' g can be interpreted like Cohen's d : A Hedges' g of $< |0.2|$ denotes no effect, $\geq |0.2| < |0.5|$ a small effect, $\geq |0.5| < |0.8|$ a moderate effect, and $\geq |0.8|$ a large effect (Cohen, 1988). A positive g indicates that human-likeness leads to a positive social response. The signs of effect sizes were adapted, so that positive effect sizes consistently stand for beneficial outcomes. I could calculate many effect sizes based on the information on means, standard deviations, and group sizes provided in the papers. Other frequently reported information included unstandardized or standardized regression coefficients, t , η^2_{partial} , and F values.

⁹ Two of the three manuscripts that were unpublished at the time of analysis were published afterwards.

Table 9 Overview of Moderator Variables

Variable	Definition
Sample	
Age	M_{age} of participants in a sample
Gender	% of female participants in a sample
Experience	% of participants in a sample who have used a chatbot before
Agent	
Assigned gender	Whether the assigned gender is female (vs. non-female)
Visual representation	Whether the avatar is human-like (vs. not human-like)
Interaction mechanism	Whether the interaction is unstructured (vs. structured)
Manipulated social cue	Whether the manipulated social cue is verbal (vs. other)
Technology	Whether the chatbot is rule-based (vs. other)
Context	
Type	Whether the application context is hedonic (vs. functional)
Task criticality	Whether the task is critical (vs. non-critical)
Method	
Sample	Whether a sample contains only students (vs. non-students)
Study setting	Whether a study was conducted online (vs. lab/field)
Stimulus type	Whether a stimulus consists of an interaction (vs. prerecorded)
Scientific field	Whether a paper is from communication literature (vs. other)
Year of publication	Year a paper was published
Sample country	Whether a sample is US-based (vs. other country)
Publication status	Whether a paper is published (vs. not published)
Peer review	Whether a paper is peer-reviewed (vs. not peer-reviewed)
Preregistration	Whether a study was preregistered (vs. not preregistered)

Effect sizes from comparisons using the same control group were averaged to avoid double counting of participants (Higgins et al., 2024, section 23.3.4 How to include multiple groups from one study); comparisons of mean values from factorial trials were included separately (Higgins et al., 2024, section 23.3.6 Factorial trials). I calculated 800 effect sizes from 199 datasets, mostly automatedly, with the help of an R script roughly following the approach by Roesler et al. (2021). The majority of dataset samples consisted of non-students (vs. students) (70.1%), directly interacted with a chatbot as stimulus (vs. were exposed to prerecorded materials) (64.8%) and participated in an online (vs. lab or field) experiment (84.3%). Many samples stemmed from the USA (45%), followed by China (16.8%), Germany (13.4%), the Netherlands (8.1%), and others (16.8%). The datasets comprised data from 41,642 participants. Participants were, on average, 31.4 years old; around half of them identified themselves as female (50.8%). Most participants indicated prior experience with chatbots (70%). The minority of datasets were collected in the context of a preregistered study (4.5%).

I assessed risk of bias using Nudelman & Otto (2020)'s Risk of Bias Utilized for Surveys Tool (ROBUST)¹⁰. The tool consists of eight questions for eight topics answered by the author for each dataset, e.g., whether the sample size was sufficient, whether the paper provided the sample's basic demographic information, and whether measurements were reliable. A risk of bias score was created by counting the „no“ responses; higher numbers thus indicated a higher risk of bias ($M = 2.79$, $Mdn = 3$, range = 1–7).

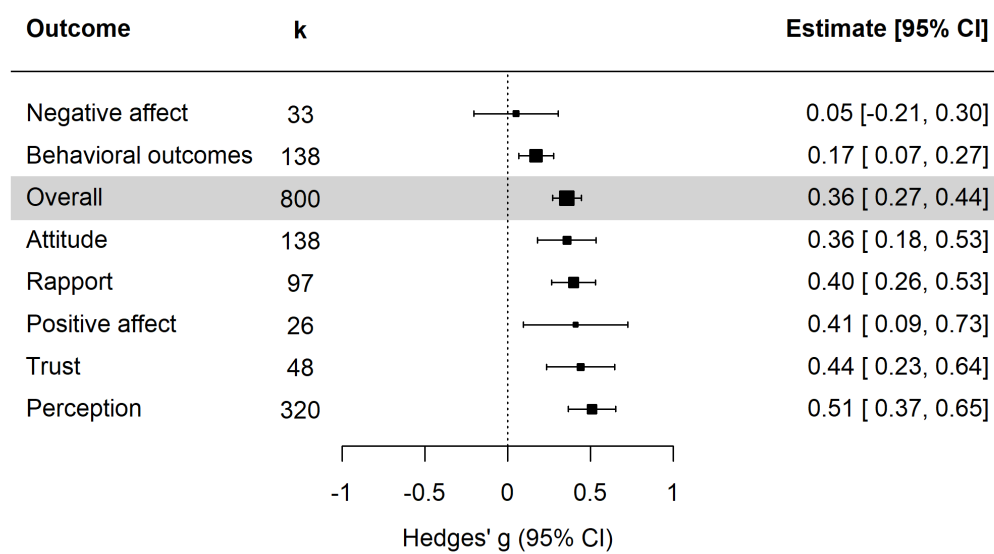
4.5 Results

4.5.1 Main Effects

To answer the first research question, I computed meta-analytic random-effects models for the human-likeness effects on the outcome variables. The data were analyzed using R (version 4.3.1) (R Core Team, 2023) and the *metafor* package (version 4.6-0) (Viechtbauer, 2010). I chose random-effects models as the dataset sample showed differences in employed methods and sample characteristics, which may lead to heterogeneity among the true effects. Heterogeneity was estimated using restricted maximum-likelihood estimation (REML) (Viechtbauer, 2010). To use all available effect size information and to account for potential non-independence induced by multiple effect sizes calculated on the same data and multiple datasets reported in the same paper, I conducted four-level meta-analyses (Hansen et al., 2022; Moeyaert et al., 2017). Before computing the models, I performed a systematic search for influential cases (Viechtbauer, 2010). The search yielded potentially influential effect sizes in all outcome categories, which I inspected in the dataset and the papers. As I found no obvious errors or misbehavior, I included all effect sizes in the analyses ($N = 800$).

Figure 11 gives an overview of the results. The detailed results of the meta-analyses and the assessment of heterogeneity are given in Table 10. The analyses yielded a small positive overall human-likeness effect on users' social responses. Human-like chatbot characteristics had a significant medium-sized positive effect on perception. Significant small positive effects emerged on attitude, rapport, positive affect, and trust. A very small significant positive effect was observed on behavioral outcomes. There was no significant effect for negative affect.

¹⁰ This systematic risk of bias assessment was not preregistered.

Figure 11 Summary Forest Plot

Note. k = number of effect sizes. Estimate = Hedges' g . 95% CI = 95% confidence interval. The size of the squares is proportional to the precision of the estimates.

Table 10 Meta-Analytic Results and Heterogeneity Assessment

Outcome	Effect sizes				Heterogeneity		
	k	g	SE	95% CI	Q	I^2	95% PI
Overall	800	0.36	.04	[0.27, 0.44]	10,131.30	95.57	[-1.00, 1.71]
Attitude	138	0.36	.09	[0.18, 0.53]	1,635.41	95.11	[-0.95, 1.66]
Perception	320	0.51	.07	[0.37, 0.65]	5,106.52	97.37	[-1.24, 2.26]
Positive affect	26	0.41	.15	[0.09, 0.73]	367.49	93.45	[-0.92, 1.02]
Negative affect	33	0.05	.12	[-0.21, 0.30]	138.88	89.58	[-1.33, 1.27]
Rapport	97	0.40	.07	[0.26, 0.53]	903.12	91.81	[-0.61, 1.40]
Trust	48	0.44	.10	[0.23, 0.64]	592.23	94.44	[-0.88, 1.75]
Behavioral outcomes	138	0.17	.05	[0.07, 0.27]	1,077.06	90.74	[-0.76, 1.10]

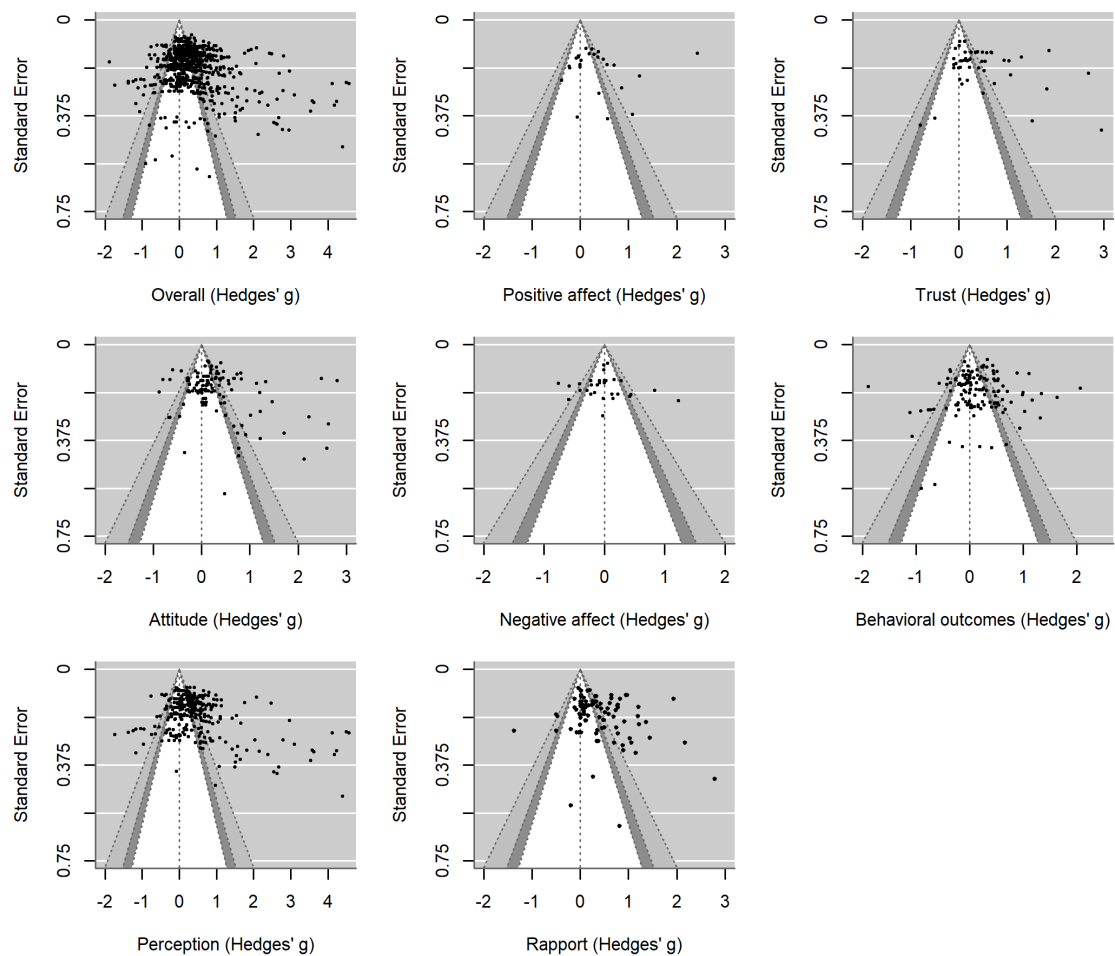
Note. Effect sizes Hedges' g and Q statistics in bold are significant ($p < .05$). k = number of effect sizes, SE and 95% CI = standard error and 95% confidence interval of effect size, Q = inverse-variance weighted sum of squared deviations between observed effects and summary effect, I^2 = the fraction of variance due to unaccounted factors, 95% PI = the range in which the true effect size lies in about 95% of all populations (Borenstein, 2019).

A substantial amount of heterogeneity was present, as indicated by the significant Q tests for all outcome categories. Except for negative affect, I^2 values indicated that over 90% of the total variability in each category was due to the influence of factors not considered in the respective model. The prediction intervals also supported heterogeneity in outcomes. Across outcome categories, they indicated that human-like characteristics can have minor, moderate, and large true effects. Although the average results were estimated to be positive, the true outcome in some datasets may be negative.

4.5.2 Publication Bias

Although I took measures to counter publication bias, e.g., by sending out a call for unpublished research, the non-publication of studies with unfavorable or nonsignificant results might have affected the results of this meta-analysis (Vevea et al., 2019). Thus, I performed Egger's regression tests for funnel plot asymmetry (Egger et al., 1997; Rodgers & Pustejovsky, 2021) and visual inspection of contour-enhanced funnel plots that indicate publication bias due to the exclusion of nonsignificant results (Peters et al., 2008). Egger's regression test indicated asymmetric funnel plots for the overall model ($p = .001$), attitude ($p = .030$), and perception ($p < .001$) but not for positive affect ($p = .944$), negative affect ($p = .080$), rapport ($p = .080$), trust ($p = .355$), and behavioral outcomes ($p = .862$). Visual inspection of the overall funnel plot suggested slight rightward bias, in other words, the estimates appeared to be rather over- than underestimated (Figure 12).

Figure 12 Contour-Enhanced Funnel Plots



Note. Plots for overall social responses and per subcategory, reference line = 0, white background: $p > .10$, dark grey background: $.05 < p < .10$, light grey background: $.001 < p < .05$, region outside of funnel: $p < .001$.

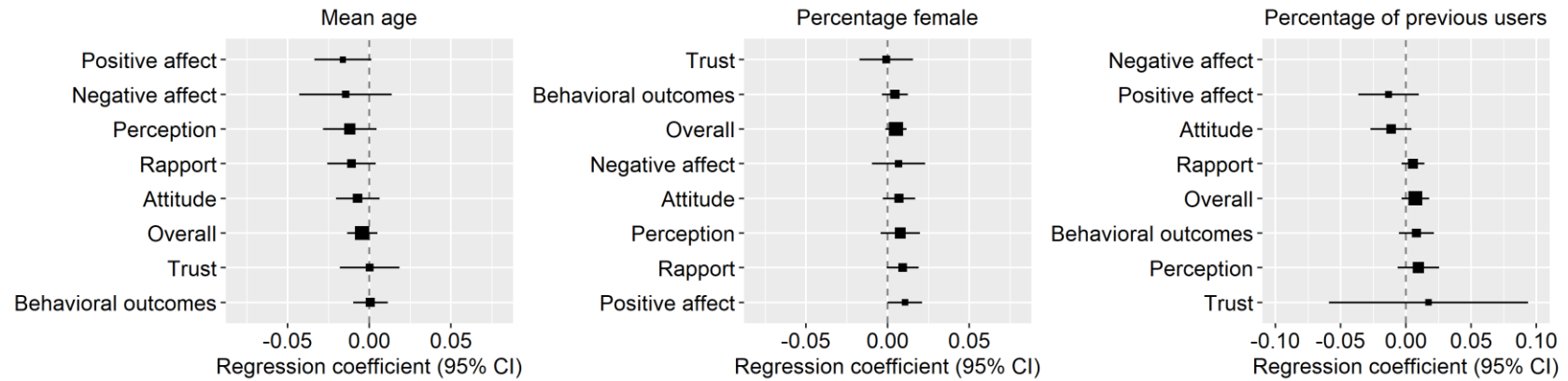
4.5.3 Moderator Effects

Subgroup analyses. Due to the substantial heterogeneity detected (Table 10), I investigated the extent to which selected sample-, agent-, context- and method-related factors influenced the main effect estimates. I conducted moderator analyses using multilevel mixed-effects models, with heterogeneity estimated via REML (Viechtbauer, 2010). First, the average effect sizes for each moderator level were estimated. Next, a test of moderators was used to check if the effect sizes differ significantly. A significant F test indicates significant differences in the effect sizes between the moderator levels. Figures 13–16 contain visual overviews of the moderator results. I present the results of the subgroup analyses for the complete sample ($N = 800$) in Table 11. The full subgroup analyses results for each outcome category can be found in Appendix C, Tables C.4–C.10.

No significant moderator effects of sample age, sample gender, and previous chatbot use were present. Regarding the agent-related moderators, the interaction mechanism and the type of social cues the chatbot possessed moderated the human-likeness effect: The overall effect was significant when the interaction mechanism was unstructured. Similar effects were found for attitude, $F(1, 51) = 10.18, p = .002, b = 0.54$, perception, $F(1, 133) = 6.19, p = .014, b = 0.38$, positive affect, $F(1, 17) = 4.55, p = .048, b = 0.61$, and trust, $F(1, 26) = 4.65, p = .040, b = 0.43$. A larger overall effect was found when human-likeness was implemented by verbal (vs. other) cues. A similar effect emerged in the models for attitude, $F(1, 132) = 5.10, p = .026, b = 0.20$, positive affect, $F(1, 23) = 8.67, p = .007, b = 0.38$, rapport, $F(1, 93) = 13.89, p < .001, b = 0.32$, and behavioral outcomes, $F(1, 129) = 5.47, p = .021, b = 0.24$. Assigned chatbot gender, avatar, and technology did not moderate human-likeness effects.

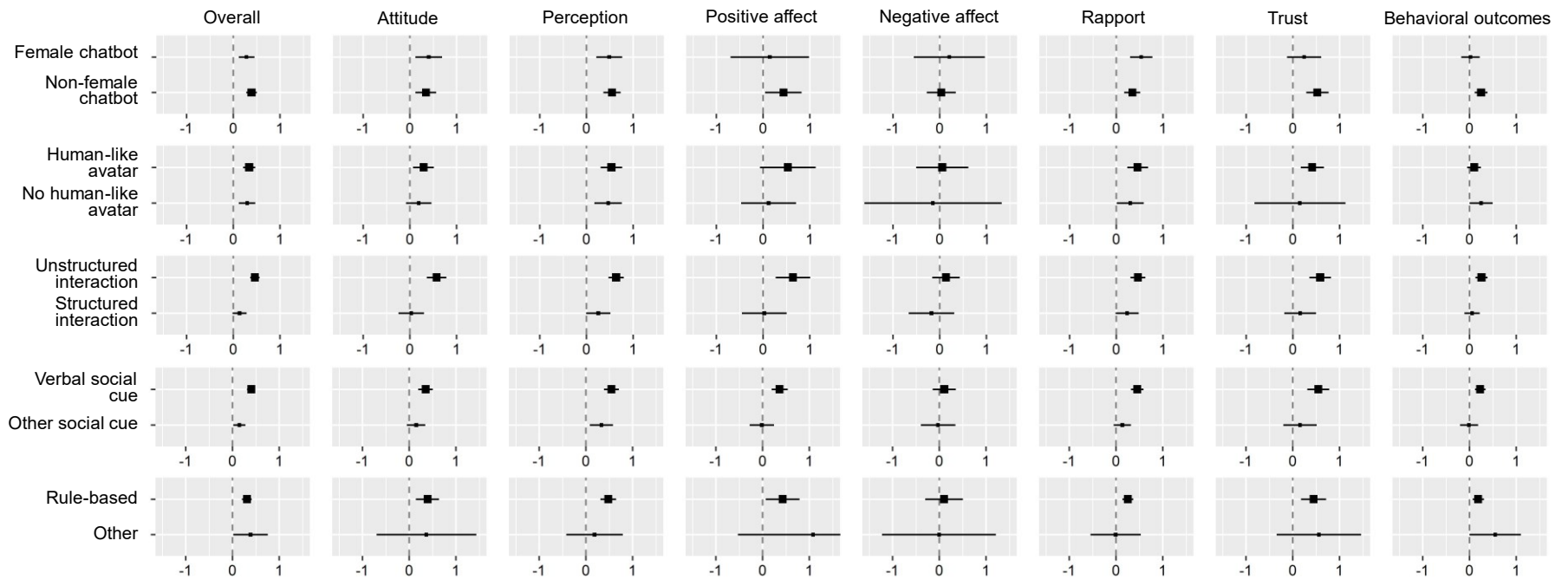
Application context did not significantly moderate the overall effect but task criticality did: The effect was larger in settings where the chatbot performed non-critical tasks. A similar but only marginally significant influence emerged on the effect on trust, with the effect being marginally significant in non-critical contexts, $F(1, 26) = 3.03, p = .093, b = 0.38$. Context type merely had a marginally significant influence on the human-likeness effect on positive affect, with the effect being significant in hedonic contexts, $F(1, 17) = 3.56, p = .076, b = 0.55$.

The only method-related moderator that significantly influenced the overall human-likeness effect was the scientific field; the effect was larger when the article was not located in the field of communication. While no other method-related moderators influenced the overall effect, some played a role in terms of positive and negative affect, and rapport. Sample type significantly moderated the effect on positive affect, $F(1, 20) = 4.42, p = .048, b = 0.31$, with the effect being larger in student samples.

Figure 13 Influence of Sample-Related Moderators on Human-Likeness Effects

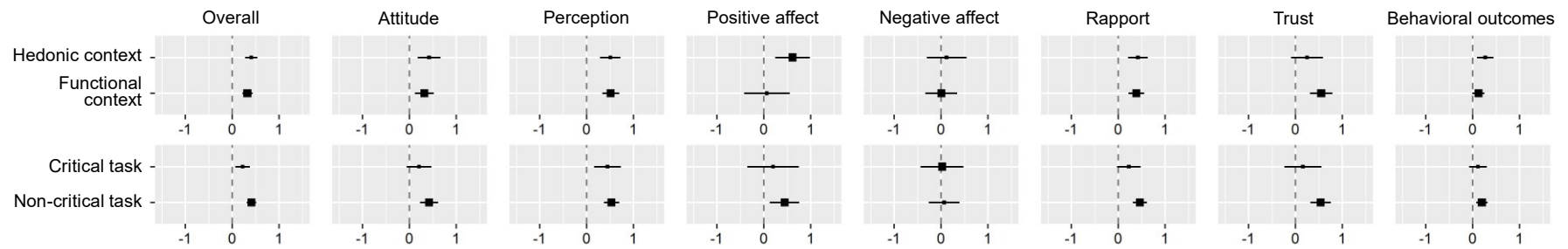
Note. The plots show the regression coefficients and the corresponding 95% confidence intervals for the influence of sample-related moderators on the overall and individual social responses. The size of the squares is proportional to the number of included effect sizes. No data were available to compute the moderating effect of prior use experience on negative affect.

Figure 14 Influence of Agent-Related Moderators on Human-Likeness Effects



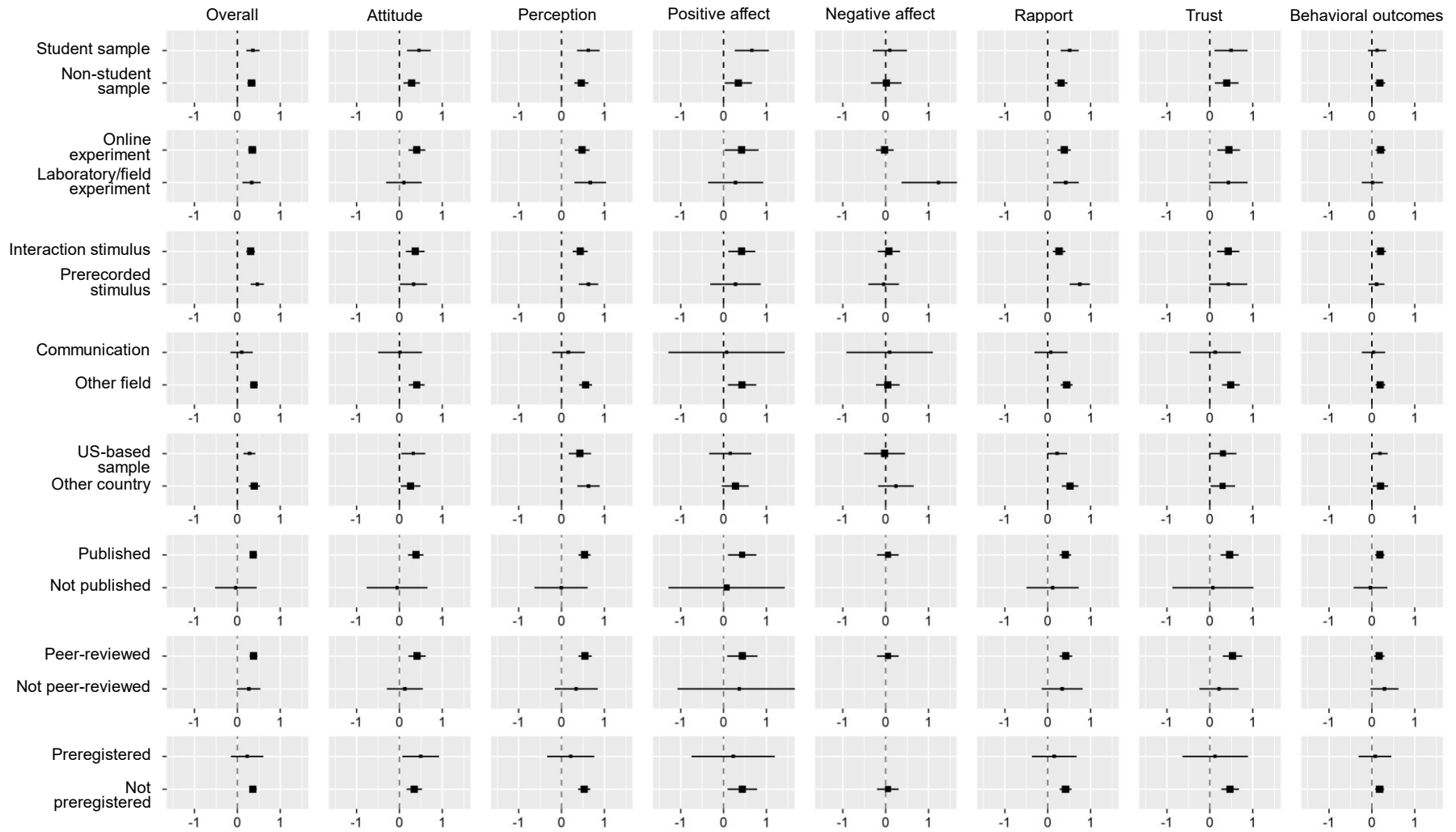
Note. The plots show the human-likeness effects overall and for each social response (Hedges' g) and the corresponding 95% confidence intervals for the agent-related moderator subgroups. The size of the squares is proportional to the number of included effect sizes.

Figure 15 Influence of Context-Related Moderators on Human-Likeness Effects



Note. The plots show the human-likeness effects overall and for each social response (Hedges' g) and the corresponding 95% confidence intervals for the context-related moderator subgroups. The size of the squares is proportional to the number of included effect sizes.

Figure 16 Influence of Method-Related Moderators on Human-Likeness Effects



Note. The plots show the human-likeness effects overall and for each social response (Hedges' g) and the corresponding 95% confidence intervals for the method-related moderator subgroups. The size of the squares is proportional to the number of included effect sizes. In the "negative affect" category, all articles were published, peer-reviewed and not preregistered. The results for the continuous moderator "year of publication" are provided in Table 11.

Table 11 Subgroup Analyses for Overall Social Responses ($k = 800$)

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	Test of Moderators
Sample mean age	719	-0.004	.005	-0.01, 0.005	-0.94	.347	0.004	.005	$F(1, 166) = 0.89, p = .347$
Percentage of female	737	0.01	.003	-0.01, 0.01	1.55	.124	0.01	.003	$F(1, 175) = 2.39, p = .124$
Percentage of previous users	161	0.01	.01	-0.003, 0.02	1.40	.172	0.01	.01	$F(1, 28) = 1.96, p = .172$
Assigned chatbot gender							0.11	.10	$F(1, 176) = 1.17, p = .281$
Female	273	0.28	.08	0.11, 0.45	3.33	.001			
Non-female	470	0.39	.06	0.28, 0.50	6.91	< .001			
Chatbot avatar							0.04	.11	$F(1, 138) = 0.16, p = .692$
Human-like	419	0.34	.07	0.21, 0.47	5.10	< .001			
Not human-like	184	0.30	.09	0.12, 0.47	3.33	.001			
Chatbot interaction							0.33	.09	$F(1, 193) = 13.15, p < .001$
Unstructured	514	0.46	.05	0.36, 0.56	9.18	< .001			
Structured	273	0.13	.08	-0.02, 0.28	1.75	.083			
Chatbot social cue							0.26	.07	$F(1, 762) = 14.15, p < .001$
Verbal	580	0.40	.05	0.31, 0.49	8.82	< .001			
Other	185	0.14	.07	0.01, 0.27	2.17	.031			
Chatbot technology							0.08	.19	$F(1, 97) = 0.17, p = .682$
Rule-based	444	0.31	.05	0.20, 0.41	5.93	< .001			
Other	28	0.39	.19	0.02, 0.75	2.07	.041			
Context							0.08	.08	$F(1, 197) = 0.98, p = .322$
Hedonic	297	0.41	.07	0.27, 0.54	6.01	< .001			
Functional	503	0.32	.06	0.21, 0.43	5.82	< .001			
Task							0.19	.09	$F(1, 197) = 4.28, p = .040$
Critical	219	0.22	.08	0.07, 0.38	2.81	.006			
Non-critical	581	0.41	.05	0.31, 0.51	8.15	< .001			
Sample							0.03	.09	$F(1, 189) = 0.11, p = .737$
Student	194	0.36	.08	0.21, 0.52	4.59	< .001			
Non-student	587	0.33	.05	0.23, 0.43	6.52	< .001			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	Test of Moderators
Study setting							0.01	.12	$F(1, 197) = 0.01, p = 1.00$
Online	648	0.35	.05	0.25, 0.44	7.07	< .001			
Other	118	0.33	.11	0.12, 0.55	3.06	< .001			
Stimulus type							0.15	.09	$F(1, 197) = 2.81, p = .095$
Interaction	506	0.31	.05	0.21, 0.41	6.00	< .001			
Prerecorded	294	0.46	.08	0.31, 0.62	6.02	< .001			
Scientific field							0.29	.14	$F(1, 140) = 4.14, p = .044$
Communication	130	0.10	.13	-0.16, 0.36	0.74	.459			
Other	670	0.39	.05	0.30, 0.48	8.42	< .001			
Year of publication	716	0.01	.01	-0.03, 0.04	0.11	.745	0.01	.01	$F(1, 137) = 0.11, p = .745$
Sample country							0.11	.10	$F(1, 147) = 1.34, p = .250$
USA	285	0.28	.07	0.14, 0.42	3.98	< .001			
Other	338	0.39	.07	0.26, 0.52	5.98	< .001			
Publication status							0.41	.25	$F(1, 140) = 2.65, p = .106$
Published	716	0.37	.04	0.28, 0.46	8.33	< .001			
Not published	84	-0.04	.25	-0.52, 0.45	-0.15	.878			
Peer review							0.11	.15	$F(1, 135) = 0.56, p = .455$
Peer reviewed	635	0.37	.05	0.28, 0.54	7.90	< .001			
Not peer reviewed	78	0.27	.14	-0.01, 0.54	1.92	.057			
Preregistration							-0.13	.20	$F(1, 197) = 0.47, p = .493$
Preregistered	86	0.23	.19	-0.15, 0.61	1.19	.236			
Not preregistered	714	0.36	.04	0.27, 0.45	8.09	< .001			

Note. *k* = number of effect sizes. Figures in bold are significant ($p < .05$).

On average, the effect on negative affect was larger for laboratory or field (vs. online) studies, $F(1, 13) = 9.36, p = .009, b = 1.25$. However, only one effect size could be included for the laboratory/field category. The effect on rapport was significantly larger when the sample was not US-based, $F(1, 47) = 4.08, p = .049, b = -0.30$, and when the study employed prerecorded materials (vs. interactions), $F(1, 56) = 12.57, p < .001, b = 0.48$. Publication status, year of publication, and whether the study was peer-reviewed or preregistered did not moderate the human-likeness effects.

In an exploratory moderator analysis, I investigated whether the study's risk of bias moderated the human-likeness effects on our outcomes of interest. Risk of bias did not significantly moderate the overall effect, $F(1, 197) = 2.34, p = .128, b = 0.05$. However, it significantly positively influenced the effects on attitude, $F(1, 67) = 4.86, p = .031, b = 0.12$, positive affect, $F(1, 20) = 4.80, p = .040, b = 0.18$, and rapport, $F(1, 67) = 6.93, p = .011, b = 0.14$ (i.e., the higher the risk of bias, the larger the effects).

Meta-Regression. Category-wise mixed-effects moderator models were run, simultaneously regressing the effect sizes on multiple moderators to validate the robustness of the results (Viechtbauer, 2010). Due to sample size considerations and to keep the results comparable, I calculated models including all moderators for which significant effects emerged in the subgroup analyses, including risk of bias (Blut et al., 2021). Multicollinearity among moderator variables was assessed using variance inflation factors (VIF). The highest VIF observed was 1.36, indicating acceptable levels of multicollinearity. The results of the meta-regression models are presented in Table 12. For the overall human-likeness effect, the meta-regression model largely replicated the results from the separate analyses. The effect of human-like chatbot characteristics was larger when chatbots employed verbal (vs. other) cues and when the interaction was unstructured (vs. structured). However, the effects of task criticality and article field vanished in the meta-regression model. Regarding attitude, the moderator effect of interaction mechanism persisted while the one of social cues became non-significant. Interaction mechanism also moderated the effect on perception. The effect on positive affect was merely moderated by interaction mechanism in the meta-regression. Like in the subgroup analyses, the human-likeness effect on negative affect was significantly influenced by study setting. Regarding rapport, the manipulation of social cues and stimulus type had a significant influence; the effect of country was marginally significant. No significant moderator effects emerged for trust or behavioral outcomes, as indicated by the non-significant Tests of Moderators.

Table 12 Meta-Regression Results

Moderator	Overall ($k = 752$)		Attitude ($k = 132$)		Perception ($k = 313$)		Positive affect ($k = 25$)	
	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>
Verbal cue (other)	0.22**	.07	<i>0.15</i> [°]	.09			0.27*	.13
Unstructured interaction (structured)	0.28**	.09	.40*	.15	0.38*	.15	<i>0.27</i>	.15
Non-critical task (critical)	<i>0.11</i>	.09						
US-based sample (other)							<i>0.14</i>	.14
Student sample (non-student)								
Communication (other)	<i>0.16</i>	.13						
Online study (lab/field)								
Interaction (prerecorded stimulus)								
Study risk of bias	0.02	.03	<i>0.07</i>	.05	0.07	.05	<i>0.09</i>	.05
Test of Moderators	$F(5, 126) = 6.39, p < .001$		$F(3, 47) = 5.50, p = .003$		$F(2, 132) = 4.05, p = .020$		$F(4, 13) = 5.31, p = .009$	

Moderator	Negative affect ($k = 33$)		Rapport ($k = 85$)		Trust ($k = 48$)		Behavioral outcomes ($k = 131$)	
	<i>b</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>
Verbal cue (other)			0.28**	.09			0.25*	.10
Unstructured interaction (structured)					0.43*	.20		
Non-critical task (critical)								
US-based sample (other)			<i>-0.23</i> [°]	.13				
Student sample (non-student)								
Communication (other)								
Online study (lab/field)	-1.18*	.43						
Interaction (prerecorded stimulus)			-0.53***	.14				
Study risk of bias	0.04	.06	0.13*	.05	0.02	.07	-0.01	.04
Test of Moderators	$F(2, 12) = 4.66, p = .032$		$F(4, 42) = 9.14, p < .001$		$F(2, 25) = 2.32, p = .119$		$F(2, 73) = 2.78, p = .069$	

Note. Due to sample size considerations, only the moderators significant in the subgroup analyses and risk of bias were included as moderators. Compared to the subgroup analyses, the effects in bold are still significant, while the effects in italics are no longer significant. *b* = unstandardized regression coefficient, *SE* = standard error.

[°] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

4.6 Discussion

This meta-analysis investigated the effects of human-likeness in text-based conversational agents on a broad range of social responses, as well as the influence of moderators related to sample, agent, context, and method. Considering 800 effect sizes from 199 experimental datasets, the analyses yielded a small positive overall effect of chatbots' human-like social cues on users' social responses. This effect varied depending on the social response category and specific moderators, in particular, whether a chatbot used verbal (vs. other) social cues to convey human-likeness, whether the interaction was unstructured (vs. structured), and whether the task performed by the chatbot was non-critical (vs. critical).

The overall finding aligns with previous meta-analyses on the impact of anthropomorphism in intelligent media agents on human-related outcomes (Blut et al., 2021; Roesler et al., 2021). However, in contrast to the moderate overall human-likeness effect for social robots found by Roesler et al. (2021) and the strong correlation between perceived anthropomorphism and intention to use service robots found by Blut et al. (2021), the effect for text-based conversational agents appears to be somewhat smaller. This might be because variations in text-based chatbots' social cues might not be as salient to users as those of more embodied agents that can leverage the potential of voice and movement. For instance, in their meta-analysis on the effectiveness of social cues in human-robot interaction, Xu et al. (2023) showed that facial, kinetic, and haptic cues have a greater influence on social presence and trust as perceived by users than verbal cues.

The size of the positive effect varied for the different outcome categories, being moderate for perception, small for outcomes related to attitude, rapport, positive affect, and trust, very small for behavioral outcomes, and not significant for negative affect. This suggests that social cues are quite effective in improving perceptual, affective, and trust-related outcomes but not as effective in improving how people (intend to) act and in improving their negative emotions. The relative difference between the relatively large human-likeness effects on perceptual, affective, and trust-related outcomes and the small human-likeness effect on behavior aligns with prior meta-analytic research on AI-based agents (G. Huang & Wang, 2023; Roesler et al., 2021). Perceptions of the agent and the interaction, trust, and positive feelings appear to be elicited quite easily using social cues. That might also be the reason why the perception category comprises the largest number of effect sizes. Behavioral intentions and actual behavior appear harder to change with human-likeness. Factors connected to the functioning of the chatbot might have superseded the effects of social cues to increase behavioral intentions and change behavior. The findings further suggest that human-likeness is

not as effective in mitigating negative emotions like frustration and stress than has been suggested by the literature (Benner et al., 2021; De Sá Siqueira et al., 2023). There is first evidence that human-likeness might increase frustration in erroneous chatbot interactions (Brendel et al., 2023), and decrease users' evaluation of a firm and their purchase intention when customers are already angry when contacting a chatbot (Crolig et al., 2022). However, this finding should be further investigated by extending the research to other negative emotions such as stress and feelings of eeriness, as well as the factors that determine the relationship between human-likeness and negative affect.

The moderator analyses revealed several interesting findings. Surprisingly, but in line with prior meta-analyses (Blut et al., 2021; G. Huang & Wang, 2023), sample characteristics did not influence the human-likeness effects on our outcomes of interest.

Strikingly, gender-specific chatbot cues did not moderate human-likeness effects either. This contrasts with findings from previous meta-analyses suggesting that human-likeness is more effective when text-based chatbots with female gender cues are employed (Blut et al., 2021). As female gender-specific cues in chatbots are very common and this meta-analysis includes more recent papers (the median year of publication was 2022), users might have gotten used to them, and thus, human-likeness effects might be similar for chatbots with non-female gender cues. Future research might disentangle gender effects by answering how participants who identify as female (non-binary, transgender, male) respond to chatbots with female, male, or ambiguous gender-specific cues. The chatbot's task should also be considered, as it might influence gender-specific stereotypes. Many chatbots in the included studies did not have gender-specific cues. However, previous research has shown that people still attribute gender, even without gender-specific cues (Abercrombie et al., 2023).

Human-likeness effects in studies that equipped their human-like chatbot with only human-like visual, invisible, or interaction cues were, on average, smaller than effects from studies that manipulated human-likeness using verbal cues. This trend also emerged for attitude, positive affect, rapport, and behavioral outcomes. The potential of social cues appears to be more fully utilized in interactions where the user communicates with the chatbot via free text (overall, and in the categories attitude, perception, positive affect, and trust). A reason might be that free text interaction comes more naturally to participants because it is more similar to a chat interaction with a human than the interaction via clicking on predefined answer options (Jain et al., 2018).

The results further suggest that task criticality (i.e., whether stakes for users are high) matters. Specifically, the overall human-likeness effect was larger for uncritical tasks (i.e.,

when stakes were low). This aligns with findings from meta-analyses on more embodied agents that concluded that perceived anthropomorphism played a larger role in noncritical services (Blut et al., 2021) and that human-likeness was more effective in social contexts, including entertainment but also therapy and education (Roesler et al., 2021).

The sample's composition appeared to influence effect sizes; specifically, student samples tended to yield larger effects (in the positive affect category) because they are likely more homogenous regarding socio-economic background, resulting in less variance than more diverse general population samples (Orsingher et al., 2010). The effect on rapport was larger in studies that used prerecorded materials. This finding is a bit puzzling but might be explained by the advantage of vignette studies that participants are able to inspect the stimulus materials more closely (Abendschein et al., 2021).

4.6.1 Implications for Theory and Research

This meta-analysis contributes to the research field of human-machine communication (HMC), particularly the literature around human-likeness, CASA, and the emerging stream on the MASA paradigm. The findings support the CASA and MASA paradigms, suggesting that social cues elicit social responses (Lombard & Xu, 2021; Nass & Moon, 2000). By showing that verbal cues in text-based chatbots are at least as effective in eliciting social responses compared to other types of social cues, this current study challenges the proposition that verbal cues are secondary social cues and per se less effective in eliciting social responses (Lombard & Xu, 2021; Xu, Chen, et al., 2023). Future experimental studies should systematically answer whether a hierarchy of social cues exists that applies to all media agents alike.

The study also revealed that social cues do not improve all social responses to the same extent as the discrepancy between moderate effects on perceptions and very small effects on behavioral outcomes illustrates. This finding challenges the CASA and MASA paradigms by signaling that “not all social responses are indicative of mindlessness” (E.-J. Lee, 2024, p. 2). While for attitude, perception, trust, positive affect, and rapport, the idea of mindless responses is supported, conscious, systematic cognitive processes might be activated when it comes to users' intention to use a chatbot. Future research should avoid seeing social responses as a single entity and instead systematically investigate differences in the effect of a social cue on different social responses (E.-J. Lee, 2024). Also, perceptions of the agent and the interaction, trust, and positive feelings toward the agent have been shown to mediate the effects of human-likeness on behavioral outcomes in previous research (Gopinath & Kasilingam, 2023). The analyses to investigate questions like these (e.g., using structural equation modeling (SEM)) were unfortunately not in the scope of this meta-analysis but should be conducted in the future.

The analyses revealed important boundary conditions for CASA and MASA. The findings indicate that the interaction should not be structured and as free as possible to leverage the benefits of social cues. In addition, human-likeness might work differently for text-based chatbots where social cues concentrate on language cues. When few social cues are available to users, it becomes increasingly difficult to distinguish whether textual information is coming from an AI/an AI-based chatbot or a human (Dou et al., 2022; Gunser et al., 2022). This issue is particularly relevant at the time of writing due to the introduction of LLM-based open-domain chatbots like OpenAI's ChatGPT (OpenAI, 2024a). It is therefore important to systematically research the negative consequences of chatbot anthropomorphism on societal issues, e.g., the spread of misinformation and disinformation and the replication of biases and stereotypes (Abercrombie et al., 2023).

This meta-analysis also uncovered which social responses are often investigated in research on human-like chatbots. The number of effect sizes per outcome category varied: Most effect sizes emerged for perception, including scales from the Godspeed questionnaire, and rapport, including scales to assess social presence. Rapport, trust, and affective outcomes comprised the fewest effect sizes. The behavioral outcomes category also included behavioral intentions, which is probably why the number of effect sizes is comparable to the attitude category. Scales assessing self-reports are generally easier to research than actual behavior (Roesler et al., 2021). However, they are also susceptible to biases arising from social desirability or the tendency to agree to statements (Wetzel & Greiff, 2018).

Although the study setting was not a significant moderator in the current research, there was a discrepancy between the large number of online studies and the relatively small number of laboratory and field experiments included in this meta-analysis. In the last few years, online samples have been established as a convenient and fast way of data collection, which was likely fueled by the COVID-19 pandemic. I recommend that researchers go into the field again and investigate the impact of real chatbots on real users. Finally, only a minority of studies were preregistered (4.5%), indicating room for improvement in open science practices.

4.6.2 Practical Implications

The findings indicate that social cues improve social responses overall. This is especially relevant for perceptions, trust, positive affect, and rapport (i.e., user-chatbot relationship outcomes). It is not as relevant for changing users' behaviors and behavioral intentions as well as for alleviating users' negative emotions. In light of effectiveness, practitioners should, therefore, keep implementing human-like chatbots. Yet, the results of the moderator analyses have important implications for chatbot designers. Employing visual,

identity, or other non-verbal cues alone might not be as effective for improving user responses—the chatbot needs to communicate in a human-like manner. Adding human-like verbal cues might lead to larger effects of social cues, confirming the current practice. The results suggest that practitioners can better leverage the potential of human-like features when the chatbot’s interaction mechanism is free text. However, previous research has shown that a chatbot needs to function reliably to fully realize the benefits of a more human-like chatbot interaction (Klein et al., 2024). With LLM-supported chatbots that function smoothly, human-like interactions could become possible like never before. The findings also indicate that human-likeness can be applied more effectively in chatbots that perform non-critical tasks where users are only exposed to low risk.

It is important to note that human-like design does not always align with users’ expectations; for example, human-likeness might signal that the chatbot possesses competencies it does not actually have (Shanahan, 2024). Some researchers have thus recommended to avoid using social cues that signal, e.g., personality, altogether (Abercrombie et al., 2023). In addition, as this meta-analysis has shown, human-likeness can lead to increased trust in or credibility of information. Prior research has shown that information provided by conversational agents like chatbots is perceived as more credible than information coming from a web page, regardless of whether the information is accurate or not (Anderl et al., 2024). Especially for text-based chatbots, where social cues are restricted to verbal and other, more subtle cues, concerns related to anthropomorphism in dialogue systems emerge because users could more easily believe they are interacting with another human. Thus, human-like chatbots can lead users to be willing to share their (personal) data with chatbots and, thus, with the individuals and organizations behind the chatbots (Singh et al., 2023). These risks have to be considered and addressed when designing chatbots. Clear and transparent design, e.g., through labels and explanations and focusing on improving users’ AI literacy, might help reduce and adjust overly high expectations and unintended data exposure (Sundar & Liao, 2023).

4.6.3 Limitations and Future Research

This meta-analysis has several limitations. Due to language restrictions, I only searched for papers in English and German. However, I am confident that this limitation does not alter the overall conclusions of this meta-analysis. Although selected moderators could partly explain the variance in effect sizes, heterogeneity was generally present in the analyses. Potential reasons for this include grouping various (single and combinations of) social cues that can signal different things (e.g., empathy, humor, personality, responsiveness) into one outcome variable and combining several constructs in one subcategory. In addition, crucial HMC

constructs like perceived human-likeness and social presence have often been operationalized differently (see Ischen et al., 2023 for an overview of human-likeness measures).

Certain factors related to risk of bias limit the validity of the study's conclusions: About 7% of eligible papers did not report sufficient information to calculate effect sizes and had to be excluded beforehand, which could have reduced the final set of articles and effect sizes. Risk of bias, as assessed through the ROBUST criteria (Nudelman & Otto, 2020) was only a significant moderator in the rapport category, as shown by the results of the meta-regressions. Furthermore, neither publication status nor whether it was a peer-reviewed publication—two other indicators of a study's quality—moderated the effects of human likeness.

I also want to point out that meta-analytic moderator analyses are generally based on observational data (i.e., article- and dataset-related characteristics), and, therefore, do not allow conclusions about causality (Viechtbauer, 2007). Although moderators were carefully selected based on theory and previous empirical research, unobserved factors might have been confounded with one or several moderators and might have distorted the findings. As the number of effect sizes was low in some categories (e.g., positive and negative affect and trust), either because the dependent variable was not investigated often or because the respective moderator information was not reported in the individual paper, statistical power might not have been sufficient to detect significant moderator effects in some cases. Additionally, I could not perform some planned moderator analyses due to insufficient data.

The stimulus chatbots in the included studies were merely text-based. However, recent technological advances have produced various other agents with which users interact in their everyday lives, e.g., voice-based assistants and virtual reality avatars (Lu et al., 2024). Whether and how human-likeness works differently in voice-based or virtual agents that do not mainly communicate via text but at the same time are not as embodied as physical robots has yet to be explored. To this end, future studies should systematically compare text-based agents and other types of agents concerning the effectiveness of social cues.

4.7 Conclusion

This meta-analysis identified a small positive effect of human-likeness on user responses, indicating that social cues are generally effective in improving social responses. However, text-based chatbots' social cues appear less effective than in more embodied agents. Consistent with previous findings, human-likeness effects varied, showing a moderate impact on perception, a small impact on attitude, rapport, positive affect, and trust, a very small impact on behavioral outcomes, and no impact on negative affect. The results also suggest that human-likeness is more effective in unstructured chatbot interactions, when the chatbot communicates

in a human-like way, and when the chatbot performs a non-critical task. The results challenge MASA's proposition that secondary cues, such as language use, are less effective in eliciting social responses than primary cues, such as a human-like picture. As human-like text-based chatbots can lead to greater self-disclosure and trust in information, practical implications for designing effective and ethically justifiable chatbot interactions emerge.

5. General Discussion

With this dissertation, I aim to improve our understanding of users' cognitive, affective, and behavioral social responses to text-based chatbots. First, this dissertation took a step back and compared users' responses to chatbots and human agents. In the next step, it zoomed in on chatbots and explored the impact of human-likeness as implemented through social cues on users' responses in four experiments. The focus here was on responsiveness, social error responses, dynamic response times, and free text interaction; previously less explored but potentially relevant verbal and interaction social cues. Finally, a meta-analysis provided generalizing insights into the effect of social cues on social responses towards text-based chatbots, taking into account possible moderators. Theoretically, the research questions were based on the Media Are Social Actors (MASA) paradigm (Lombard & Xu, 2021) and the machine heuristic of the MAIN model (Sundar, 2008; Sundar et al., 2019), established research frameworks that emphasize the impact of heuristics triggered by social cues.

Chapter 2 focused on the influence of agent type, responsive verbal cues, and dynamically delayed responses on users' cognitive, affective, and behavioral social responses toward chat advisors. The pooled results of three vignette experiments in a setting requiring human skills (i.e., developmental study advising) demonstrated that participants slightly preferred to interact with a human advisor over a chatbot, both in terms of intention to use and perceived competence. However, participants perceived the chatbot as more enjoyable. While the agent's response time did not impact users' responses, responsive verbal cues increased perceptions of likability and warmth and users' overall satisfaction with the chatbot's performance. These positive responsiveness effects were consistently mediated by social presence. In summary, Chapter 2 suggests that while users enjoy interacting with chatbots more, they have reservations about using them for a *human* task and that even subtle verbal cues can contribute positively to user satisfaction by evoking feelings of social presence.

Chapter 3 emphasized the influence of verbal social cues, i.e., social error response, and interaction cues, i.e., free text interaction, on social responses. We carried out an online experiment in which participants interacted with a customer service chatbot that was programmed to occasionally fail to understand users' requests. Contrary to our expectations, users preferred interacting with a chatbot via buttons rather than free text. This was reflected in higher perceptions of warmth and competence, satisfaction with the agent, behavioral intention, and emotional connection with the company. The structured interaction via buttons was favored, possibly due to lower perceived ease of use in the free text condition. As expected, user satisfaction increased when the chatbot responded socially to an error, i.e., when it

apologized to the user for not understanding a query. The significant indirect effects of free text interaction via perceived warmth and competence indicated that these concepts are crucial for people's evaluations of new actors. In sum, Chapter 3 corroborated the importance of a smooth chatbot interaction for users' reuse intent and emotional connection with the company while demonstrating again that subtle verbal cues effectively increase user satisfaction.

Chapter 4 took a meta-perspective and looked at the effect of human-likeness on seven categories of social responses towards chatbots and potential moderating factors related to features of the sample, the agent, the context, and the method. Due to conflicting conclusions of previous research and because prior meta-analyses mostly looked at embodied robots (e.g., Roesler et al., 2021), the HMC literature still lacks a comprehensive meta-analysis on the impact of chatbots' human-likeness on social responses (J. Kim et al., 2024). To address this gap, I conducted a PRISMA-based meta-analysis (Page et al., 2021). As expected, human-like chatbot characteristics exerted a small positive effect on social responses. Yet, while the human-likeness effect on behavioral outcomes was very small (and non-existent on negative affect), the effects on perceptions, attitude, positive affect, and rapport were small to moderate. Moderator analyses showed that the human-likeness effect was larger when human-likeness was implemented via verbal social cues (i.e., when the chatbot communicated in a human-like way) than via other (e.g., visual, invisible) cues. In contrast to Chapter 3, human-likeness was more effective when unstructured interactions via free text were possible as well as in non-critical contexts. Several method-related moderators played a role, e.g., human-likeness effects were larger when studies employed prerecorded stimulus material (vs. actual interactions; for rapport), researched students (for positive affect), were conducted in the lab or field (vs. online; for negative affect), and when the sample was not from the US (for rapport). Taken together, Chapter 4 highlights that human-likeness improves social responses effectively, but to varying degrees, and that agent-, context-, and method-related features moderate these effects.

5.1 Main Conclusions and Theoretical Implications

Four overarching conclusions emerge from the findings of Chapters 2–4: The first conclusion refers to people's different responses to chatbots and human agents; the second conclusion regards the potential of human-likeness to improve social responses; the third conclusion relates to the effectiveness of human-likeness for different social responses and the fourth conclusion focuses on boundary conditions of the effectiveness of human-likeness. In the following section, I interpret each conclusion, situate it within the literature, and draw implications for theory and research.

First, the findings presented in this dissertation suggest that mindful processes may play a larger role in evaluating an agent than previously thought. In Chapter 2, users were more likely to consult the human advisor again and evaluated her as more competent than the chatbot. These findings support the reverse machine heuristic, which states that machine-like interface cues have a deterrent effect, and users, therefore, prefer human agents to execute *human* tasks like study advising (Sundar et al., 2019). The prerecorded interactions in Chapter 2 were smooth, ended with an appropriate recommendation, and differed only in the agent's identity. Still, a human agent was preferred, indicating an ongoing aversion to the use of automation and algorithms in tasks traditionally attributed to humans (Dietvorst et al., 2015). Another relevant factor may be users' source orientation (Van der Goot, 2022). Prior research has shown that in their interactions with agents, users do not always think they are simply interacting with a chatbot or a human agent (Van der Goot, 2022). According to the Source Orientation Model (Solomon & Wash, 2014), users can perceive different layers behind a technology that can influence their cognitive, affective, and behavioral responses. In the case of chatbots, users might orient themselves toward the programmer who developed the chatbot, the organization that employs the chatbot, or even the technology provider (Van der Goot, 2022). Future research could experimentally investigate which levels users perceive and whether the context or the task that an agent performs and its degree of human-likeness, influence this perception. The findings have implications for the MASA paradigm and the idea that people's responses to media agents are mindless (Lombard & Xu, 2021). Recent research suggests mindful responses to technology may depend on situational factors like the application context (E.-J. Lee, 2024). People might be mindful of the source if they "have a reason to consider who authored a news article, who checked the facts, who composed a music, and the like" (E.-J. Lee, 2024, p. 6). Similarly, when users get advice regarding their path of study, which can lead to an impactful decision, they might think twice about whether they prefer to interact with a machine-like chatbot or an empathetic and caring human being.

Yet, not all outcomes appear to be equally affected. Although, as expected, participants' intention to use a means of communication and their competence perception of the advisor were slightly higher for the human agent, the effects were very small. Moreover, participants indicated slightly higher levels of perceived enjoyment for the chatbot. In addition, there were no differences in overall attitude, perceptions of likability, intelligence and warmth, and satisfaction with the agent's performance. The overall means of all outcomes were generally high, and the (non-existent) effects of agent type did not change, regardless of whether the agent was equipped with responsive verbal cues. This is consistent with the conclusion of a recent

meta-analysis by G. Huang and Wang (2023) that diverse AI agents (i.e., in customer service, automated journalism, and content moderation) and humans elicit similar responses regarding overall evaluations and perceptions of the agent, in turn confirming the MASA paradigm that states that people perceive media agents as social actors and therefore react to them socially (Lombard & Xu, 2021). However, the prerecorded conversations shown to participants in the vignettes in Chapter 2 were relatively smooth compared to interactions with actual customer service chatbots (e.g., the one participants interacted with in Chapter 3). The differences could have been more pronounced if the chatbot had not understood the user's input. Still, with LLM-powered chatbots enabling smoother and more human-like interactions, the vignettes may now be closer to reality than when the experiments were conducted.

Taken together, although users appear to treat chatbots and human agents similarly, differences seem to remain, even when both agents perform equally well. Further research should focus on the discrepancy between attitudes/perceptions and behavioral intentions, as multiple mechanisms, including the reverse machine heuristic (Sundar, 2020), the concept of algorithm aversion (Dietvorst et al., 2015), and different source orientations (Solomon & Wash, 2014) may play a role (G. Huang & Wang, 2023; E.-J. Lee, 2024).

Second, the findings of this dissertation demonstrate that relatively subtle verbal cues can lead users to perceive chatbots as communicative subjects and improve social responses. The results of the experimental studies in Chapter 2 already suggested that verbal social cues like responsive backchanneling and social error responses have the potential to impact social responses positively. In contrast, dynamically delayed response times (Chapter 2) and free text interaction (Chapter 3) did have no or reverse effects on social responses, demonstrating that not all social cues effectively improve social responses. The meta-analysis presented in Chapter 4 eventually confirmed the difference in the effectiveness between verbal and other social cues for text-based chatbots. The findings challenge MASA's theoretical assumption that secondary cues (e.g., language use) are "less salient and less central to humans' perception of socialness" (Lombard & Xu, 2021, p. 32f.) and thus less effective in eliciting social responses than primary cues (e.g., visual representations). This finding might be because verbal cues matter more for chatbots as text is the mode of communication here (Araujo, 2018), and cues that affect the communication mode itself might be more salient than other cues (e.g., invisible or interaction cues).

Furthermore, this dissertation enhances the understanding of the mechanisms underlying the effects of social cues on social responses. I looked at established and newer concepts relating to perceptions of the chatbot (i.e., perceived warmth and competence) and the

interaction experience (i.e., social presence, perceived dialogue, feeling heard) that social cues might signal to users. Social presence has been confirmed (e.g., Go & Sundar, 2019) as a mediator of the effect of responsive verbal cues on more social-oriented agent perceptions (i.e., likability and warmth) and, more importantly, on users' satisfaction with the interaction (Chapter 2). Perceptions of a dialogue and users' feeling of being heard by the agent also partially informed this effect. Empirical evidence for agent perceptions of warmth and competence as explanatory mechanisms of the relationship between interaction mechanism and satisfaction, behavioral intention, and company connection was found in Chapter 3. The findings show that several mechanisms can explain why people respond to chatbots socially.

Researchers could strive to systematically test whether a “hierarchy of social cues that represents their different power in evoking users' social responses” (Xu, Liu, et al., 2023, p. 344) exists, which cues make up this hierarchy, and whether and how this hierarchy applies to different kinds of media agents, including chatbots, voice-based CAs, virtual reality avatars, and embodied robots (Xu, Chen, et al., 2023). Plus, the power of combinations of social cues (e.g., as implemented by Schanke et al. (2021)) as well as individual user and context differences to predict social responses should be investigated to achieve a complete picture of why people respond to media agents the way they do (Xu, Liu, et al., 2023).

Third, this dissertation concludes that human-likeness does not improve all social responses to the same extent. This dissertation assessed a comprehensive range of social response outcome measures, including attitudinal, affective, and behavioral measures. In Chapters 2–4, human-likeness influenced different social responses to varying degrees. The experiments conducted in two different service contexts yielded small effects of verbal social cues on likability and warmth perceptions and overall satisfaction, but the impact did not extend to general attitudes, behavioral intentions, and emotional connection with the company. The pattern found in the experimental studies was also present in the meta-analytic results. Small to moderate human-likeness effects on perception, attitude, rapport, positive affect, and trust contrasted with a very small effect on behavioral outcomes and no significant effect on negative affect. Human-likeness, thus, seems to be not as effective in changing behavioral intentions and behaviors as well as in lowering people's negative emotions. An explanation could lie in the way we see social responses. The concept is often used as an umbrella term for people's cognitive, affective, and behavioral reactions towards media agents like chatbots (Krämer, 2005; E.-J. Lee, 2024). Social responses are often treated as a single entity without considering that “not all social responses are equally indicative of mindlessness” (E.-J. Lee, 2024, p. 2). Consequently, not all social responses might result from triggered heuristics (Lombard & Xu,

2021); some might result from a more mindful, systematic processing approach. So, at least when comparing two chatbot versions, the more human-like version might not always automatically elicit more positive social responses. This aligns with previous research on the detrimental effects of human-likeness in chatbots, which suggests that human-likeness may not be beneficial in all situations, e.g., when people experience negative feelings like anger (Crolic et al., 2022) or when a chatbot gives incorrect answers (Brendel et al., 2023). As suggested by the relatively low number of effect sizes in the negative affect category in Chapter 4, experimental research that systematically investigates the negative consequences of human-likeness is necessary.

Fourth, this dissertation extends current knowledge about the boundary conditions of the effectiveness of social cues. While sample characteristics like age and prior chatbot experience did not exert significant influence on the human-likeness effects in Chapter 4, selected characteristics of the agent, the context, and the method seemed to matter. Overall, and in the subcategories attitude, perception, positive affect, and trust, effect sizes were larger when the chatbot was based on unstructured interaction via free text (vs. structured). This contrasts with the findings from Chapter 3, in which users generally preferred interacting with chatbots via predefined buttons, regardless of whether the agent was equipped with verbal cues (i.e., social responses to errors). As also shown in Chapter 3, whether social cues are more effective in unstructured free text interaction may depend on how easy users find the chatbot to use, which in turn depends on the chatbot's performance. Since the chatbot in our experiment (Chapter 3) was programmed to sporadically not understand the user's input, and users encountered more mistakes in the free text conditions, users preferred the more structured chatbot. Still, this dissertation included two different approaches to designing conversational flows as stimulus materials. In Chapter 2, the conversations shown in the vignettes were quite smooth and without interruptions but potentially trailblazing and applicable to LLM-based chatbots, whereas, in Chapter 3, the conversations were erroneous with built-in situations where the chatbot did not understand the users' request, closely resembling chatbots currently employed by firms (Chapter 3). The effectiveness of verbal cues for user satisfaction was demonstrated in both situations.

In addition, different scenarios where chatbots will potentially continue to be relevant in the future were investigated. In contrast to prior work that looked at the impact of agent type and social cues on user responses in task-based customer service scenarios (Go & Sundar, 2019; Lou et al., 2022; Song & Shin, 2024), Chapter 2 focused on developmental study advising, a task that needs human competencies (Chapter 2). On the contrary, the experiment in Chapter 3

took place in a more functional customer service setting. This dissertation shows that human-like verbal cues can improve user satisfaction in both contexts (Chapters 2 and 3). Chapter 4 suggests that social cues might be more effective when the chatbot executes non-critical tasks like engaging in small talk or providing recipe recommendations. Future research should test this proposition experimentally. In addition, while the experimental studies examined two different service contexts (i.e., study advising and customer service), the meta-analysis included application context as a moderator of the human-likeness effects on social responses. However, no differences between hedonic and functional contexts were found.

Future studies could look at additional moderating factors stemming from the emerging field of Explainable AI (XAI), “the class of systems that provide visibility into how an AI system makes decisions and predictions and executes its actions” (Rai, 2020, p. 137f.). Xu and Shi (2024) have recently tried to integrate HMC with XAI, proposing a model “in which users’ responses can be conceptualized as outcomes of the interactions between their reactions to machine interfaces [...] and their understanding of machines’ working mechanisms [...]” (Xu & Shi, 2024, p. 4). Specifically for text-based chatbots, the levels of the chatbot’s message personalization (standardized vs. tailored responses), message production (pattern-matching algorithms vs. unsupervised learning), and message accessibility (private vs. public) might influence the effectiveness of the social cues displayed by the chatbot (Xu & Shi, 2024).

5.2 Strengths

The conceptual framework of this dissertation consisted of two established theoretical approaches in HMC research that both stress the importance of heuristics for users’ responses towards media agents (Dehnert & Mongeau, 2022). Whereas the machine heuristic of the MAIN model is suited for comparing social responses to different information sources (i.e., human vs. chatbot), the MASA paradigm is suited for researching communication between humans and machines (Lombard & Xu, 2021; Sundar & Kim, 2019). All studies assessed a comprehensive range of social response measures, including attitudes towards, evaluations, and perceptions of the chatbot and the interaction experience, as well as affective and behavioral outcomes. Common HMC concepts, such as social presence, as well as concepts from interpersonal relations like feeling heard (Roos et al., 2023) (Chapter 2), perceived warmth and competence (Fiske, 2018), and marketing, e.g., emotional connection with the company (Christodoulides et al., 2006) (Chapter 3), were included and proved to be transferable to human-chatbot interaction. Thus, in addition to the theoretical implications outlined in the previous paragraph, this dissertation suggests that users treat media agents as communicative

subjects as opposed to mere tools (Guzman, 2018), which is in line with the chosen HMC approach, and the MASA paradigm (Lombard & Xu, 2021).

The findings of this dissertation contribute to multiple streams of literature. First, this thesis contributes to the HMC literature around source orientation, including the machine heuristic (E.-J. Lee, 2024). The findings enhance our understanding of users' social responses to chatbots and human agents by highlighting that although chatbots perform well as study advisors, users still prefer to consult a human agent.

Second, the thesis adds to the literature on the influence of social cues on social responses towards (text-based) conversational agents on two levels (Abercrombie et al., 2023; Chaves & Gerosa, 2021; Rapp et al., 2021). It provides new experimental insights on the impact of previously overlooked individual social cues and enhances our understanding of the underlying mechanisms of the effects of these social cues on social responses (Chapters 2 and 3). Moreover, from a generalizing perspective, this dissertation provides a quantitative synthesis of 800 effect sizes gathered from experimental research articles on the impact of human-likeness on social responses and an exploration into the factors that inform this relationship. The findings add to our understanding of the effects of human-likeness, and complement prior review articles on the topic, for instance, Rapp et al. (2021)'s systematic review with a quantified measure (Chapter 4). The work provides a point of reference for other researchers wishing to investigate the effects of human-likeness in other types of media agents, such as voice-based CAs or avatars in virtual environments (Lu et al., 2024).

Third, this thesis contributes to the service literature by examining relevant outcome variables like satisfaction, intention to use, and emotional connection with the company. It provides recommendations for enhancing service quality and the relationship between organizations and consumers. Plus, it sheds light on the mechanisms underlying service outcomes, highlighting that perceptions of the interaction experience (i.e., social presence and perceived dialogue; Chapter 2) and perceptions of the agent (i.e., perceived warmth and competence; Chapter 3) are central for improving satisfaction.

Methodwise, this dissertation included two different kinds of experiments, as well as a meta-analysis. In contrast to surveys, experimental studies in HMC are high in internal validity and allow for statements about causality between the manipulated independent variables and the outcome variables (Greussing et al., 2022). Vignette experiments in which participants are exposed to prerecorded materials, as used in Chapter 2, offer participants the possibility to closely read the materials without having to engage in an interaction (Abendschein et al., 2021). In the experiment in Chapter 3, participants engaged in an interactive experience with a state-

of-the-art chatbot in a realistic environment as an experimental stimulus, which increased ecological validity. The chatbot was developed with the support of a conversational AI company that could provide us with real-life experience and expert knowledge in creating chatbots for customer service. The meta-analysis presented in Chapter 4 was based on a comprehensive literature search for published and unpublished work and followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA; Page et al., 2021), ensuring transparent and reproducible reporting of methods and results.

5.3 Methodological Limitations

This dissertation has limitations regarding the methods applied that have to be kept in mind and could be addressed in future research. While vignette designs, as used in Chapter 2, exhibit high internal validity, they might have decreased external validity as participants were mere observers of the interaction with the chat agent. To counter this limitation, Chapter 3 employed a more externally valid design by letting participants engage with a customer service chatbot to solve pre-defined tasks. The meta-analysis in Chapter 4 included many studies in which participants engaged in real interaction with a chatbot, but field studies were rare. Although sometimes difficult to achieve, researchers should aim for external validity by seeking collaborations with industry partners to research human-chatbot interactions in the field.

Furthermore, the experiments considered different use cases (i.e., study advising at a university and customer service at an online retailer) with specific stimulus materials, raising the question of whether the findings apply to other settings. Although we tried to make the scenarios as stereotypical and natural as possible (Jackson & Jacobs, 1983) by researching current chatbot applications in study advising (Chapter 2) and by collaborating with an industry partner (Chapter 3), the results could be due to the characteristics of the materials used. Future research should thus test our assumptions with alternative stimuli in different contexts.

While the samples in the experimental studies were based in Germany, the meta-analysis included samples from worldwide. However, even in the meta-analysis, samples from the “Western world” were in the majority. A recent analysis of social media discussions on the public perception of AI agents revealed differences in perceptions between participants from the US and China (Z. Liu et al., 2024). Whereas AI agents were functional tools for participants from the US, Chinese participants saw a more hedonic potential. Although the subgroup analyses in Chapter 4 did not yield differences between samples from the US and other countries (except a smaller human-likeness effect on rapport for US samples), future research would do well in exploring cultural differences more closely.

Another methodological limitation relates to conclusions about causality. The mediation analyses in Chapters 2 and 3, specifically the relationships between the mediators and the social responses, and the moderator analyses in Chapter 4 were based on cross-sectional, observational data, meaning no conclusions can be drawn about causality. I propose conducting experimental studies in which the results of the moderator analyses (e.g., concerning different application contexts) could be systematically tested.

Finally, all experiments assessed a broad range of self-reported social response measures without including direct behavioral measures like self-disclosure or purchase decisions. Also, in light of the relatively small meta-analytic human-likeness effect on behavioral outcomes, future studies could shed light on the impact of social cues on users' behavioral responses to chatbots. In addition, future research should examine the impact of social cues in chatbots on economically relevant outcomes beyond user satisfaction and intention to use, like purchase intention, brand image, and firm performance.

5.4 Practical Implications

Implications arise for organizations, particularly for decision-makers and developers, and for other societal actors, especially for educational institutions and policymakers. The results of this dissertation can help improve the strategic use and design of chatbots in organizations. This thesis revealed that while attitudes and perceptions of chatbots and human advisors are similar, there is still a difference in people's intention to use them and in their competence perception (Chapter 2). Decision-makers should, therefore, carefully assess which tasks require a chatbot and when human employees are more appropriate. For example, in tasks that require high levels of intuitive and empathetic intelligence, such as complaint management, humans might be more capable of providing understanding and comfort than chatbots (M.-H. Huang & Rust, 2018). While many organizations are trying to automate consumer processes with the help of AI (Costello & LoDolce, 2022), it could be useful to ask consumers directly in which situations and for which tasks they expect or prefer advice from a human or a chatbot.

Although this dissertation has shown that consumers seem to have reservations about using chatbots for some tasks, organizations are increasingly enthusiastic about using AI and chatbots for their purposes—a development set to continue (Costello & LoDolce, 2022). Decision-makers implementing a chatbot application must ensure enough funds and skilled staff to build and maintain the application for smooth, accurate functioning (Bishop et al., 2024). Even chatbots powered by LLMs require careful management to consistently provide accurate and reliable answers (L. Chen et al., 2024). Humanizing a chatbot with verbal cues (e.g., responsiveness and apology) can be done quickly and effectively (Chapters 2–4). Still,

developers and decision-makers must remember that these cues do not impact all social responses to the same extent; for instance, user satisfaction, perceptions, and emotions might be more affected than behavioral outcomes (Chapters 2–4), which might raise challenges when it comes to selling a product or a service.

While this dissertation has shown that human-like social cues can effectively improve cognitive, affective, and behavioral social responses to service chatbots, we should also consider the potential detrimental consequences of humanizing chatbots; in other words: Is more human-likeness always desirable? Developers should be mindful of users' tendency to personify chatbots by trying to manage expectations and avoid creating the illusion of overly human-like capabilities (Abercrombie et al., 2023; Shanahan, 2024), for example, by not using a picture of a human to represent the chatbot. Also, if the chatbot appears overly human-like or almost lifelike, it might fall into the *uncanny valley* and elicit feelings of eeriness and revulsion (Mori et al., 2012). This concept has received most attention in human-robot interaction but is increasingly applied to text-based chatbots (e.g., Ciechanowski et al., 2019; Song & Shin, 2024). Further research is needed to fully understand the concept in connection with chatbots.

As demonstrated in Chapter 4, human-likeness increases trust outcomes such as information credibility. Previous research has shown that information from a human-like conversational agent is perceived as more credible than information provided by a website, regardless of whether the information is accurate (Anderl et al., 2024). This can encourage the spread of mis- and disinformation (Abercrombie et al., 2023; Knight & Rogers, 2024), one of the greatest risks of our time (World Economic Forum, 2024). Further, users may form emotional bonds with humanized chatbots that may (negatively) affect users' relationships with other humans, for example, because people have developed scripts specific to their interactions with media agents, which they apply in their interactions with other humans (Gambino et al., 2020; Knight & Rogers, 2024). Regulations have recently been put in place to increase the transparency for users interacting with human-like AI systems. For example, the European Union's AI law states that AI systems must be appropriately designed so that users know that they are interacting with such a system and not with another human being (Regulation (EU) 2024/1689 Laying down Harmonised Rules on Artificial Intelligence, 2024). But especially in light of recent advances in the personalization of LLMs, more legislation is needed to protect users from an increased tendency to anthropomorphize LLM-based chatbots and thus disclose personal or sensitive information or become too attached (Kirk et al., 2024). One way to achieve more transparency, educate users, and create trust when designing chatbots is by implementing XAI (Rai, 2020; Xu & Shi, 2024). Further, by promoting people's knowledge of how AI-based

chatbots work, what they can do, and the psychology behind them, educational institutions and policymakers might counteract the potentially harmful effects of (over)anthropomorphizing chatbots.

5.5 Outlook and Future Research Questions

The field of research in which this dissertation is located is in constant flux. Therefore, questions regarding the applicability and generalizability of the findings to different current and future media agents arise. This dissertation focused on text-based conversational agents. However, speech technology has greatly evolved recently, leading to highly capable voice-based conversational agents. Recent research has shown that voice-based interfaces may lead to more positive service experiences and firm outcomes (Zierau et al., 2023). Still, situations in which text-based chatbots exhibit benefits over voice-based agents are likely to stay. For instance, Schindler et al. (2024) showed that consumers' satisfaction is higher when they interact with text-based (vs. voice-based) chatbots over perceived utilitarian (vs. hedonic) products because writing leads consumers to argue based on reason. Additional boundary conditions of the effectiveness of voice-based interactions lie in the level of semantic disfluency and the number of conversational turns—both may decrease service experiences and behavioral outcomes (Zierau et al., 2023). In addition, personal preferences resulting from individual psychological dispositions and psychological needs can promote a need for personalization of modality, e.g., introverted individuals might prefer interacting via text. Still, prior research has also shown that a text-based chatbot yielded more dialogue turns, longer interactions, lower mental load (Le Bigot et al., 2004), increased processing fluency, and more positive consumer responses (Schwede et al., 2022). The positive impact of text-based chatbots on processing fluency could have implications for contexts that require understanding and processing complex information, e.g., in information search or education (Anderl et al., 2024).

Another emergent and potentially highly influential technology that will shape the future of human-like text-based chatbots refers to LLMs. LLM-powered chatbots are highly capable and able to communicate with people in a natural, human-like manner, almost undistinguishable from real human interlocutors; in other words, they are “a patchwork of less-than-human with superhuman capacities, of uncannily human-like with peculiarly inhuman behaviors” (Shanahan, 2024, p. 79). OpenAI's newest LLM GPT-4o can show a personality, express emotions, and even sarcasm (OpenAI, 2024b). Regarding AI in the workplace, this is likely only the beginning of a process that will either lead to humans and AI complementing each other and working together (best-case scenario) or AI entirely replacing human workers (worst-case scenario) (M.-H. Huang & Rust, 2018). The fear that machines will replace human jobs

has existed for as long as there have been chatbots. Today, we are closer than ever to what M.-H. Huang and Rust (2021) call “feeling AI” (p. 32). While the lower end of feeling AI has already been reached with voice assistants such as Siri and chatbots for customer service, the upper end of feeling AI, capable of performing intuitive and empathetic tasks previously mostly performed by humans, has yet to develop its full potential (M.-H. Huang & Rust, 2021a). Not only will service chatbots employed by organizations likely be more human-like and personalized in the future, but other groups of society, for example, politicians, pop stars, and billionaires, might also exploit the benefits of personalized AI by communicating with and influencing their audience on a more personal level (Kirk et al., 2024). Therefore, AI-powered chatbots (especially in combination with other forms of generative AI producing images, audio, or video) can truly have transformative impacts.

This dissertation showed that humanizing chatbots via social cues effectively improves social responses to service chatbots. Regarding the future of HMC research, calls have come up to overcome the “anthropocentric bias” (Gambino et al., 2020, p. 79) inherent in our idea of how media agents should look and behave. Rather, with increasing exposure to, experience with, and knowledge about media agents, people are likely to have developed specific scripts for their interactions with them (Gambino et al., 2020). Moving HMC research further in this direction could prove fruitful, for example, to investigate the situations in which people would rather consult a chatbot or other media agent than a human. How these HMC scripts manifest, what triggers them, and how they affect social responses are questions that need to be answered by future research.

Not least because we may very soon live “in a world where some of [our] conversational partners we’ll know to be humans, some we’ll know to be bots, and probably some we won’t know either way, and may not even care” (Dale, 2016, p. 815), new research questions arise that should be approached in the future: Will humans still be able to discern a chatbot from a human (and vice versa)? How can we ensure that everyone using a conversational agent understands what is happening? How much human-likeness is enough, and can human-like media agents become an obstacle to successful communication with them and other people? Are people more or less likely to apply scripts specific to interpersonal communication in future interactions with machines, or have they developed specific scripts for HMC (Gambino et al., 2020)? What are the consequences of human-like conversational agents for everyday life and society? These questions are all relevant to the HMC research field. However, because they potentially affect multiple populations, and answering them may require additional, non-quantitative research methods (Greussing et al., 2022; J. Kim et al., 2024; Van der Goot &

Etzrodt, 2023), this dissertation calls for a greater emphasis on interdisciplinary research efforts and diverse research methods.

5.6 Conclusion

People increasingly interact with AI-based service chatbots in their everyday lives, e.g., when ordering food, looking for a new car, or selecting a new insurance company. However, these service interactions do not always turn out to be satisfactory. This dissertation aimed to provide a better understanding of users' social responses to service chatbots. First, I investigated if users prefer a service interaction with a chatbot or a human study advisor. Next, I focused on chatbots in two service scenarios (i.e., study advising and customer service). Experimental findings suggest that people prefer interacting with a human agent in a task requiring human competencies. Four experiments and a meta-analysis demonstrated that human-likeness improves cognitive, affective, and behavioral responses to varying degrees and under certain circumstances. Perceptions of the interaction experience (i.e., social presence, perceived dialogue, and feeling heard) mediated the positive effects of human-like verbal cues on user satisfaction.

This dissertation lays the foundation for future research on the influence of human-like social cues on users' social responses to text-based chatbots and other conversational agents and the potential beneficial and detrimental consequences. Given that increasingly powerful technologies such as LLMs might shortly render chatbots indistinguishable from humans, researching the impact of agent identity and human-likeness on social responses remains crucial. With the advent of LLM-powered conversational agents and inherent risks, such as the increased spread of misinformation due to hallucinations, privacy concerns, and the perpetuation of biases, there will certainly be no shortage of challenges to tackle in the future.

References

- Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(7), 16021. <https://doi.org/10.2196/16021>
- Abendschein, B., Edwards, C., & Edwards, A. (2021). The influence of agent and message type on perceptions of social support in human-machine communication. *Communication Research Reports*, *38*(5), 304–314. <https://doi.org/10.1080/08824096.2021.1966405>
- Abercrombie, G., Cercas Curry, A., Dinkar, T., Rieser, V., & Talat, Z. (2023). Mirages. On anthropomorphism in dialogue systems. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 4776–4790). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.290>
- Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2), 427–445. <https://doi.org/10.1007/s12525-020-00414-7>
- Allouch, M., Azaria, A., & Azoulay, R. (2021). Conversational agents: Goals, technologies, vision and challenges. *Sensors*, *21*(24), 8448. <https://doi.org/10.3390/s21248448>
- Anderl, C., Klein, S. H., Sarigül, B., Schneider, F. M., Han, J., Fiedler, P. L., & Utz, S. (2024). Conversational presentation mode increases credibility judgements during information search with ChatGPT. *Scientific Reports*, *14*(1), 17127. <https://doi.org/10.1038/s41598-024-67829-6>
- Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189. <https://doi.org/10.1016/j.chb.2018.03.051>
- Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 254). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300484>
- Auspurg, K., Hinz, T., & Liebig, S. (2009). Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey. *Methoden, Daten, Analysen*, *3*(1), 59–96. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-126659>

- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1, 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., Antunes, R., Gomes, M., Gatti, C., Vanzin, M., Junior, S. C., Silva, E., & Moreira, C. (2020). Conversational agents in business: A systematic literature review and future research directions. *Computer Science Review*, 36, 100239. <https://doi.org/10.1016/j.cosrev.2020.100239>
- Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, 71(3), 409–427. <https://doi.org/10.1080/10510974.2020.1725082>
- Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? A review on recovery strategies to overcome conversational breakdowns of conversational agents. In *International Conference on Information Systems (ICIS)*. https://aisel.aisnet.org/icis2021/hci_robot/hci_robot/13
- Bergner, A. S., Hildebrand, C., & Häubl, G. (2023). Machine talk: How verbal embodiment in conversational AI shapes consumer–brand relationships. *Journal of Consumer Research*, 50(4), 742–764. <https://doi.org/10.1093/jcr/ucad014>
- Bing (Producer). (2023, February 7). *The new Bing | Spark creativity* [Video recording]. https://www.youtube.com/watch?v=UHMb5vSP_U8
- Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, 12(5), 456–480. <https://doi.org/10.1162/105474603322761270>
- Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass, O. (2016). What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior*, 63, 416–423. <https://doi.org/10.1016/j.chb.2016.05.064>
- Bishop, E., Brackenbury, J., & Dixon, J. (2024, April 9). *Customer service leaders must invest in AI to enable, not replace, employees. Q&A with Emily Potosky*. Gartner.

- <https://www.gartner.com/en/newsroom/press-releases/2024-04-09-customer-service-leaders-must-invest-in-ai-to-enable-not-replace-employees>
- Blut, M., Wang, C., Wunderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*, 632–658. <https://doi.org/10.1007/s11747-020-00762-y>
- Borenstein, M., & Hedges, L. V. (2019). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd edition, pp. 207–243). Russell Sage Foundation.
- Borji, A. (2023). *A categorical archive of ChatGPT failures*. arXiv. <https://doi.org/10.48550/arXiv.2302.03494>
- Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In Y. Kompatsiaris (Ed.), *Internet Science. INSCI 2017 Lecture Notes in Computer Science* (Vol. 10673, pp. 377–392). Springer. https://doi.org/10.1007/978-3-319-70284-1_30
- Brendel, A. B., Hildebrandt, F., Dennis, A. R., & Riquel, J. (2023). The paradoxical role of humanness in aggression toward conversational agents. *Journal of Management Information Systems*, *40*(3), 883–913. <https://doi.org/10.1080/07421222.2023.2229127>
- Burgoon, J. K., & Hale, J. L. (1988). Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs*, *55*(1), 58–79. <https://doi.org/10.1080/03637758809376158>
- Cameron, D., Saille, S., Collins, E. C., Aitken, J. M., Cheung, H., Chua, A., Loh, E. J., & Law, J. (2021). The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in Human Behavior*, *114*, 106561. <https://doi.org/10.1016/j.chb.2020.106561>
- Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, *90*, 315–330. <https://doi.org/10.1016/j.chb.2018.08.048>
- Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human-chatbot interaction design. *International Journal of Human-Computer Interaction*, *37*(8), 729–758. <https://doi.org/10.1080/10447318.2020.1841438>
- Chen, J., Guo, F., Ren, Z., Li, M., & Ham, J. (2023). Effects of anthropomorphic design cues of chatbots on users' perception and visual behaviors. *International Journal of Human-*

- Computer Interaction*, 40(14), 3636–3654.
<https://doi.org/10.1080/10447318.2023.2193514>
- Chen, L., Zaharia, M., & Zou, J. (2024). How is ChatGPT's behavior changing over time? *Harvard Data Science Review*, 6(2). <https://doi.org/10.1162/99608f92.5317da47>
- Cheng, Y., & Jiang, H. (2020). How do AI-driven chatbots impact user experience? Examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use. *Journal of Broadcasting & Electronic Media*, 64(4), 592–614.
<https://doi.org/10.1080/08838151.2020.1834296>
- Christodoulides, G., Chernatony, L., Furrer, O., Shiu, E., & Abimbola, T. (2006). Conceptualising and measuring the equity of online brands. *Journal of Marketing Management*, 22(7–8), 799–825. <https://doi.org/10.1362/026725706778612149>
- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539–548.
<https://doi.org/10.1016/j.future.2018.01.055>
- Clustaar. (2019, January 1). 10 Best customer support chatbots. *Chatbots Life*.
<https://chatbotslife.com/10-best-customer-support-chatbots-5be630b7b081>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
<https://doi.org/10.4324/9780203771587>
- Costello, K., & LoDolce, M. (2022, July 27). *Gartner predicts chatbots will become a primary customer service channel within five years: Chatbot investment on the rise but low ROI and other challenges persist*. Gartner. <https://www.gartner.com/en/newsroom/press-releases/2022-07-27-gartner-predicts-chatbots-will-become-a-primary-customer-service-channel-within-five-years>
- Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the bot: Anthropomorphism and anger in customer–chatbot interactions. *Journal of Marketing Research*, 86(1), 132–148. <https://doi.org/10.1177/00222429211045687>
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811–817.
<https://doi.org/10.1017/S1351324916000243>
- Davis, D. (1982). Determinants of responsiveness in dyadic interaction. In W. Ickes & E. S. Knowles (Eds.), *Personality, Roles, and Social Behavior* (pp. 85–139). Springer.
https://doi.org/10.1007/978-1-4613-9469-3_4
- Davis, D., & Perkowski, W. T. (1979). Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on

- interpersonal attraction. *Journal of Personality and Social Psychology*, 37(4), 534–550. <https://doi.org/10.1037/0022-3514.37.4.534>
- De Andrés-Sánchez, J., & Gené-Albesa, J. (2024). Assessing attitude and behavioral intention toward chatbots in an insurance setting: A mixed method approach. *International Journal of Human–Computer Interaction*, 40(17), 4918–4933. <https://doi.org/10.1080/10447318.2023.2227833>
- De Ruyter, K., & Wetzels, M. G. M. (2000). The impact of perceived listening behavior in voice-to-voice service encounters. *Journal of Service Research*, 2(3), 276–284. <https://doi.org/10.1177/109467050023005>
- De Sá Siqueira, M. A., Müller, B. C. N., & Bosse, T. (2023). When do we accept mistakes from chatbots? The impact of human-like communication on user experience in chatbots that make mistakes. *International Journal of Human–Computer Interaction*, 40(11), 2862–2872. <https://doi.org/10.1080/10447318.2023.2175158>
- Dehnert, M., & Mongeau, P. A. (2022). Persuasion in the age of artificial intelligence (AI): Theories and complications of AI-based persuasion. *Human Communication Research*, 48(3), 386–403. <https://doi.org/10.1093/hcr/hqac006>
- Desai, R. (2023, October 17). Streamlining customer support with ChatGPT chatbots! *Medium*. <https://medium.com/@rutujadesai/streamlining-customer-support-with-chatgpt-chatbots-d26913a1a826>
- Diederich, S., Brendel, A. B., Lichtenberg, S., & and Kolbe, L. (2019). Design for fast request fulfillment or natural interaction? Insights from an experiment with a conversational agent. *Proceedings of the 27th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/20
- Diers, T. (2020). *Akzeptanz von Chatbots im Consumer-Marketing: Erfolgsfaktoren zwischen Konsumenten und künstlicher Intelligenz*. Springer Gabler. <https://doi.org/10.1007/978-3-658-29317-8>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Dixon, M., V. Karniouchina, E., van der Rhee, B., Verma, R., & Victorino, L. (2014). The role of coordinated marketing-operations strategy in services: Implications for managerial decisions and execution. *Journal of Service Management*, 25(2), 275–294. <https://doi.org/10.1108/JOSM-02-2014-0060>

- Dou, Y., Forbes, M., Koncel-Kedziorski, R., Smith, N. A., & Choi, Y. (2022). Is GPT-3 text indistinguishable from human text? Scarecrow: A Framework for Scrutinizing Machine Text. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 7250–7274). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.501>
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)*, *315*(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138–161. <https://doi.org/10.1016/j.ijhcs.2019.07.009>
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender bias in chatbot design. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science* (Vol. 11970, pp. 79–93). Springer. https://doi.org/10.1007/978-3-030-39540-7_6
- Ferraro, C., Demsar, V., Sands, S., Restrepo, M., & Campbell, C. (2024). The paradoxes of generative AI-enabled customer service: A guide for managers. *Business Horizons*, *67*(5), 549–559. <https://doi.org/10.1016/j.bushor.2024.04.013>
- Fiore, S. M., Wiltshire, T. J., Lobato, E. J. C., Jentsch, F. G., Huang, W. H., & Axelrod, B. (2013). Toward understanding social cues and signals in human-robot interaction: Effects of robot gaze and proxemic behavior. *Frontiers in Psychology*, *4*(859), 2–15. <https://doi.org/10.3389/fpsyg.2013.00859>
- Fishbach, A., & Labroo, A. A. (2007). Be better or be merry: How mood affects self-control. *Journal of Personality and Social Psychology*, *93*, 158–173. <https://doi.org/10.1037/0022-3514.93.2.158>

- Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, 27(2), 67–73. <https://doi.org/10.1177/0963721417738825>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Fogg, B. J. (Ed.). (2003). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann.
- Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulou, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing*, 103, 2915–2942. <https://doi.org/10.1007/s00607-021-01016-7>
- Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. In S. S. Bodrunova (Ed.), *Internet Science. INSCI 2018. Lecture Notes in Computer Science* (Vol. 11193, pp. 194–208). Springer. https://doi.org/10.1007/978-3-030-01437-7_16
- Følstad, A., & Skjuve, M. (2019). Chatbots for customer service: User experience and motivation. In B. R. Cowan & L. Clark (Eds.), *Proceedings of the 1st International Conference on Conversational User Interfaces*. Association for Computing Machinery. <https://doi.org/10.1145/3342775.3342784>
- Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2019). Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design. In S. Bodrunova (Ed.), *Internet science. INSCI 2018. Lecture Notes in Computer Science* (Vol. 11551, pp. 145–156). https://doi.org/10.1007/978-3-030-17705-8_13
- Følstad, A., & Taylor, C. (2020). Conversational repair in chatbots for customer service: The effect of expressing uncertainty and suggesting alternatives. In A. Følstad, T. Araujo, S. Papadopoulou, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer*

- Science* (Vol. 11970, pp. 201–214). Springer. https://doi.org/10.1007/978-3-030-39540-7_14
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3), 143–166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
- Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, 1. <https://doi.org/10.30658/hmc.1.1>
- Fortunati, L., & Edwards, A. (2021). Moving ahead with human-machine communication. *Human-Machine Communication*, 2, 7–28. <https://doi.org/10.30658/hmc.2.1>
- Fotheringham, D., & Wiles, M. A. (2022). The effect of implementing chatbot customer service on stock returns: An event study analysis. *Journal of the Academy of Marketing Science*, 51, 802–822. <https://doi.org/10.1007/s11747-022-00841-2>
- Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction*, 35(6), 456–467. <https://doi.org/10.1080/10447318.2018.1456150>
- Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1, 71–86. <https://doi.org/10.30658/hmc.1.5>
- Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega*, 32(6), 407–424. <https://doi.org/10.1016/j.omega.2004.01.006>
- Gelbrich, K., Hagel, J., & Orsingher, C. (2021). Emotional support from a digital assistant in technology-mediated services: Effects on customer satisfaction and behavioral persistence. *International Journal of Research in Marketing*, 38(1), 176–193. <https://doi.org/10.1016/j.ijresmar.2020.06.004>
- Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. *Proceedings of the 26th European Conference on Information Systems (ECIS)*, 113. https://aisel.aisnet.org/ecis2018_rp/113
- Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2022). Opposing effects of response time in human–chatbot interaction. *Business & Information Systems Engineering*, 64, 773–791. <https://doi.org/10.1007/s12599-022-00755-x>

- 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>
- Gopinath, K., & Kasilingam, D. (2023). Antecedents of intention to use chatbots in service encounters: A meta-analytic review. *International Journal of Consumer Studies, 47*(6), 2367–2395. <https://doi.org/10.1111/ijcs.12933>
- Gordon, V. N. (1994). Developmental advising: The elusive ideal. *NACADA Journal, 14*(2), 72–76. <https://doi.org/10.12930/NACADA-19-201>
- Grand View Research. (2023, March 8). *Chatbot market size, share, trends & growth report, 2030*. <https://www.grandviewresearch.com/horizon/outlook/chatbot-market-size/global>
- Greussing, E., Gaiser, F., Klein, S. H., Straßmann, C., Ischen, C., Eimler, S., Frehmann, K., Giesemann, M., Knorr, C., Lermann Henestroza, A., Räder, A., & Utz, S. (2022). Researching interactions between humans and machines: Methodological challenges. *Publizistik, 67*(4), 531–554. <https://doi.org/10.1007/s11616-022-00759-3>
- Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems, 144*, 113515. <https://doi.org/10.1016/j.dss.2021.113515>
- Groom, V., Chen, J., Johnson, T., Kara, F. A., & Nass, C. (2010). Critic, compatriot, or chump?: Responses to robot blame attribution. *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 211–217. <https://doi.org/10.1109/HRI.2010.5453192>
- Grudin, J., & Jacques, R. (2019). Chatbots, humbots, and the quest for artificial general intelligence. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 209). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300439>
- Gunkel, D. J. (2020). *An introduction to communication and artificial intelligence*. polity.
- Gunser, V. E., Gottschling, S., Brucker, B., Richter, S., Çakir, D., & Gerjets, P. (2022). The pure poet: How good is the subjective credibility and stylistic quality of literary short texts written by an artificial intelligence tool as compared to texts written by human authors? *Proceedings of the Annual Meeting of the Cognitive Science Society, 44*(44). <https://escholarship.org/uc/item/1wx3983m>

- Guzman, A. L. (2018). What is human-machine communication, anyway. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology and ourselves* (pp. 1–28). Peter Lang.
- Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media & Society*, 22(1), 70–86. <https://doi.org/10.1177/1461444819858691>
- Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: A practical guide. *Management Review Quarterly*, 72(1), 1–19. <https://doi.org/10.1007/s11301-021-00247-4>
- Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). “*I think this is the most disruptive technology*”: Exploring sentiments of ChatGPT early adopters using Twitter data. arXiv. <https://doi.org/10.48550/arXiv.2212.05856>
- Haugeland, I. K. F., Følstad, A., Taylor, C., & Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. *International Journal of Human-Computer Studies*, 161, 102788. <https://doi.org/10.1016/j.ijhcs.2022.102788>
- Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). Guilford Press.
- Heyselaar, E. (2023). The CASA theory no longer applies to desktop computers. *Scientific Reports*, 13(1), 19693. <https://doi.org/10.1038/s41598-023-46527-9>
- Higgins, J. P. T., Eldridge, S., & Li, T. (Eds.). (2024). Including variants on randomized trials. In *Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024)*. Cochrane. <https://training.cochrane.org/handbook/current/chapter-23>
- Ho, S. M., Hancock, J. T., Booth, C., & Liu, X. (2016). Computer-mediated deception: Strategies revealed by language-action cues in spontaneous communication. *Journal of Management Information Systems*, 33(2), 393–420. <https://doi.org/10.1080/07421222.2016.1205924>
- Hodgins, H. S., & Liebeskind, E. (2003). Apology versus defense: Antecedents and consequences. *Journal of Experimental Social Psychology*, 39(4), 297–316. [https://doi.org/10.1016/S0022-1031\(03\)00024-6](https://doi.org/10.1016/S0022-1031(03)00024-6)
- Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & Mctear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? In M. Mulvenna (Ed.), *Proceedings of the 31st*

- European Conference on Cognitive Ergonomics* (pp. 207–214). Association for Computing Machinery. <https://doi.org/10.1145/3335082.3335094>
- Holtgraves, T. M., Ross, S. J., Weywadt, C. R., & Han, T. L. (2007). Perceiving artificial social agents. *Computers in Human Behavior*, 23(5), 2163–2174. <https://doi.org/10.1016/j.chb.2006.02.017>
- Hoxmeier, J., & DiCesare, C. (2000). System response time and user satisfaction: An Experimental Study of Browser-based Applications. *AMCIS 2000 Proceedings*, 347. <https://aisel.aisnet.org/amcis2000/347>
- Huang, G., & Wang, S. (2023). Is artificial intelligence more persuasive than humans? A meta-analysis. *Journal of Communication*, 73(6), 552–562. <https://doi.org/10.1093/joc/jqad024>
- Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Huang, M.-H., & Rust, R. T. (2021a). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. <https://doi.org/10.1007/s11747-020-00749-9>
- Huang, M.-H., & Rust, R. T. (2021b). Engaged to a robot? The role of AI in service. *Journal of Service Research*, 24(1), 30–41. <https://doi.org/10.1177/1094670520902266>
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). SAGE Publications, Ltd. <https://doi.org/10.4135/9781412985031>
- Ischen, C., Araujo, T., van Noort, G., Voorveld, H., & Smit, E. (2020). “I am here to assist you today”: The role of entity, interactivity and experiential perceptions in chatbot persuasion. *Journal of Broadcasting & Electronic Media*, 64(4), 615–639. <https://doi.org/10.1080/08838151.2020.1834297>
- Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2020). Privacy concerns in chatbot interactions. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science* (Vol. 11970, pp. 34–48). Springer. https://doi.org/10.1007/978-3-030-39540-7_3
- Ischen, C., Smit, E., & Wang, E. (2023). *Assessing human-likeness perceptions: Measurement scales of conversational agents [paper presentation]*. CONVERSATIONS 2023 – the 7th International Workshop on Chatbot Research and Design, Oslo, Finland.

- https://2023.conversations.ws/wp-content/uploads/2023/11/conversations_2023_positionpaper_13_ischen.pdf
- ISO. (2018, January 1). *Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts*. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>
- Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research*, 9(2), 169–191. <https://doi.org/10.1111/j.1468-2958.1983.tb00691.x>
- Jacobson, J., & Gorea, I. (2023). Human–machine communication in retail. In *The Sage Handbook of Human–Machine Communication* (pp. 532–539). SAGE Publications Ltd. <https://doi.org/10.4135/9781529782783>
- Jaimes, A., & Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*, 108(1–2), 116–134. <https://doi.org/10.1016/j.cviu.2006.10.019>
- Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of chatbots. *Proceedings of the 2018 Designing Interactive Systems Conference*, 895–906. <https://doi.org/10.1145/3196709.3196735>
- Jensen, K. B. (2008). Communication theory and philosophy. In *The International Encyclopedia of Communication*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405186407.wbiecc103>
- Kamoen, N., McCartan, T., & Liebrecht, C. (2022). Conversational agent voting advice applications: A comparison between a structured, semi-structured, and non-structured chatbot design for communicating with voters about political issues. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design. CONVERSATIONS 2021* (Vol. 13171, pp. 160–175). Springer. https://doi.org/10.1007/978-3-030-94890-0_10
- Kang, E., & Kang, Y. A. (2024). Counseling chatbot design: The effect of anthropomorphic chatbot characteristics on user self-disclosure and companionship. *International Journal of Human-Computer Interaction*, 40(11), 2781–2795. aph.
- Kaptein, M., Markopoulos, P., Ruyter, B., & Aarts, E. (2011). Two acts of social intelligence: The effects of mimicry and social praise on the evaluation of an artificial agent. *AI & SOCIETY*, 26(3), 261–273. <https://doi.org/10.1007/s00146-010-0304-4>
- Karrenbauer, C., König, C. M., & Breitner, M. H. (2021). Individual digital study assistant for higher education institutions: Status quo analysis and further research agenda. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation Through Information Systems*.

- WI 2021* (Vol. 48, pp. 108–124). Springer. https://doi.org/10.1007/978-3-030-86800-0_8
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review*, *28*(1), 21–37. [https://doi.org/10.1016/S0363-8111\(02\)00108-X](https://doi.org/10.1016/S0363-8111(02)00108-X)
- Kerr, E. B., & Hiltz, S. R. (1982). *Computer-mediated communication systems: Status and evaluation*. Academic Press. <https://doi.org/10.1016/C2013-0-10967-1>
- Khenouche, F., Elmir, Y., Himeur, Y., Djebari, N., & Amira, A. (2024). Revolutionizing generative pre-trained: Insights and challenges in deploying ChatGPT and generative chatbots for FAQs. *Expert Systems with Applications*, *246*, 123224. <https://doi.org/10.1016/j.eswa.2024.123224>
- Kim, J., Etzrodt, K., Prah, A., & Edwards, C. (2024). Communicating with machines: Future trajectories in the field of human-machine communication. *Global Journal of Media Studies*, *11*, 71–105. <https://doi.org/10.26599/GJMS.2024.9330006>
- Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. *Journal of Applied Psychology*, *89*(1), 104–118. <https://doi.org/10.1037/0021-9010.89.1.104>
- Kim, S., Lee, S.-Y., & Lee, J. (2021). Male, female, or robot?: Effects of task type and user gender on expected gender of chatbots. *한국멀티미디어학회논문지*, *24*(2), 320–327. <https://doi.org/10.9717/KMMS.2020.24.2.320>
- Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241–250. <https://doi.org/10.1016/j.chb.2011.09.006>
- Kirk, H. R., Vidgen, B., Röttger, P., & Hale, S. A. (2024). The benefits, risks and bounds of personalizing the alignment of large language models to individuals. *Nature Machine Intelligence*, *6*(4), 383–392. <https://doi.org/10.1038/s42256-024-00820-y>

- Klein, S. H., Papiés, D., & Utz, S. (2024). How interaction mechanism and error responses influence users' responses to customer service chatbots. *International Journal of Human-Computer Interaction*, 1–19. <https://doi.org/10.1080/10447318.2024.2351707>
- Klein, S. H., & Utz, S. (2024). Chatbot vs. human: The impact of responsive conversational features on users' responses to chat advisors. *Human-Machine Communication*, 8, 73–99. <https://doi.org/10.30658/hmc.8.4>
- Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017). The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In O. Mival, M. Smyth, & P. Dalsgaard (Eds.), *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 555–565). Association for Computing Machinery. <https://doi.org/10.1145/3064663.3064672>
- Knight, W., & Rogers, R. (2024, August 8). OpenAI warns users could become emotionally hooked on its voice mode. *Wired*. <https://www.wired.com/story/openai-voice-mode-emotional-attachment/>
- Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01086>
- Konya-Baumbach, E., Biller, M., & 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>
- Krämer, N. C. (2005). Social communicative effects of a virtual program guide. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Intelligent Virtual Agents. IVA 2005. Lecture Notes in Computer Science* (Vol. 3661, pp. 442–453). Springer. https://doi.org/10.1007/11550617_37
- Krämer, N. C., Bente, G., Eschenburg, F., & Troitzsch, H. (2009). Embodied conversational agents. *Social Psychology*, 40(1), 26–36. <https://doi.org/10.1027/1864-9335.40.1.26>
- Lagace, R. R., Dahlstrom, R., & Gassenheimer, J. B. (1991). The relevance of ethical salesperson behavior on relationship quality: The pharmaceutical industry. *Journal of Personal Selling & Sales Management*, 11(4), 39–47. <https://doi.org/10.1080/08853134.1991.10753888>
- Le Bigot, L., Jamet, E., & Rouet, J.-F. (2004). Searching information with a natural language dialogue system: A comparison of spoken vs. written modalities. *Applied Ergonomics*, 35(6), 557–564. <https://doi.org/10.1016/j.apergo.2004.06.001>

- Lee, B.-K., Lee, E. H., & Lee, T. (David). (2023). The effect of E-Government website evaluation on user satisfaction and intention to use: The mediating role of warmth and competence judgment on government. *Information, Communication & Society*, 26(9), 1868–1889. <https://doi.org/10.1080/1369118X.2022.2041701>
- Lee, E.-J. (2024). Minding the source: Toward an integrative theory of human–machine communication. *Human Communication Research*, 50(2), 184–193. <https://doi.org/10.1093/hcr/hqad034>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 205395171875668. <https://doi.org/10.1177/2053951718756684>
- Lee, M. K., Kiesler, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010). Gracefully mitigating breakdowns in robotic services. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 203–210). IEEE. <https://doi.org/10.1109/HRI.2010.5453195>
- Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, 103, 95–105. <https://doi.org/10.1016/j.ijhcs.2017.02.005>
- Lee, S., Lee, N., & Sah, Y. J. (2020). Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human-Computer Interaction*, 36(10), 930–940. <https://doi.org/10.1080/10447318.2019.1699748>
- Lenhard, W., & Lenhard, A. (2022). *Computation of effect sizes*. Psychometrica. <https://doi.org/10.13140/RG.2.2.17823.92329>
- Letheren, K., Kuhn, K.-A. L., Lings, I., & Pope, N. K. Ll. (2016). Individual difference factors related to anthropomorphic tendency. *European Journal of Marketing*, 50(5/6), 973–1002. <https://doi.org/10.1108/EJM-05-2014-0291>
- Lew, Z., & Walther, J. B. (2023). Social scripts and expectancy violations: Evaluating communication with human or AI chatbot interactants. *Media Psychology*, 26(1), 1–16. <https://doi.org/10.1080/15213269.2022.2084111>
- Liu, W., & Yao, M. Z. (2023). Human–machine communication in marketing and advertising. In *The SAGE Handbook of Human–Machine Communication* (pp. 524–531). SAGE Publications Ltd. <https://doi.org/10.4135/9781529782783>

- Liu, Z., Li, H., Chen, A., Zhang, R., & Lee, Y.-C. (2024). Understanding public perceptions of AI conversational agents: A cross-cultural analysis. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 155. <https://doi.org/10.1145/3613904.3642840>
- Liu-Thompkins, Y., Okazaki, S., & Li, H. (2022). Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *Journal of the Academy of Marketing Science*, 50(6), 1198–1218. <https://doi.org/10.1007/s11747-022-00892-5>
- LoDolce, M., & Brackenbury, J. (2023, June 15). *Gartner survey reveals only 8% of customers used a chatbot during their most recent customer service interaction*. Gartner. <https://www.gartner.com/en/newsroom/press-releases/2023-06-15-gartner-survey-reveals-only-8-percent-of-customers-used-a-chatbot-during-their-most-recent-customer-service-interaction>
- Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The Media Are Social Actors Paradigm. *Human-Machine Communication*, 2, 29–55. <https://doi.org/10.30658/hmc.2.2>
- Lou, C., Kang, H., & Tse, C. H. (2022). Bots vs. Humans: How schema congruity, contingency-based interactivity, and sympathy influence consumer perceptions and patronage intentions. *International Journal of Advertising*, 41(4), 655–684. <https://doi.org/10.1080/02650487.2021.1951510>
- Lu, X., Shi, Z., Zhang, X., Cao, L., & Liang, S. (2024). Exploring the field of virtual avatar: A bibliometric and content analysis. *International Journal of Human–Computer Interaction*, 1–17. <https://doi.org/10.1080/10447318.2024.2400405>
- Mai, V., Neef, C., & Richert, A. (2022). “Clicking vs. writing”—The impact of a chatbot’s interaction method on the working alliance in AI-based coaching. *Coaching | Theorie & Praxis*, 8(1), 15–31. <https://doi.org/10.1365/s40896-021-00063-3>
- Maio, G. R., Haddock, G., & Verplanken, B. (2019). *The psychology of attitudes and attitude change* (3rd edition). SAGE Publications Ltd.
- Maisel, N., Gable, S. L., & Strachman, A. (2008). Responsive behaviors in good times and in bad. *Personal Relationships*, 15(3), 317–338. <https://doi.org/10.1111/j.1475-6811.2008.00201.x>
- Mauldin, M. L. (1994). Chatterbots, tinymuds, and the Turing test: Entering the Loebner prize competition. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*. AAAI Press. <https://cdn.aaai.org/AAAI/1994/AAAI94-003.pdf>

- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. <https://doi.org/10.1287/isre.13.3.334.81>
- Meng, J., Rheu, M. (Mj), Zhang, Y., Dai, Y., & Peng, W. (2023). Mediated social support for distress reduction: AI chatbots vs. human. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1–25. <https://doi.org/10.1145/3579505>
- Meyer von Wolff, R., Nörtemann, J., Hobert, S., & Schumann, M. (2020). Chatbots for the information acquisition at universities – A student’s view on the application area. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science* (Vol. 11970, pp. 231–244). Springer. https://doi.org/10.1007/978-3-030-39540-7_16
- Miller, G. R. (2013). On being persuaded: Some basic distinctions. In *The SAGE handbook of persuasion: Developments in theory and practice* (2nd ed., pp. 70–82). Sage Publications.
- Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, 20(6), 559–572. <https://doi.org/10.1080/13645579.2016.1252189>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4, 1. <https://doi.org/10.1186/2046-4053-4-1>
- Moon, Y. (1999). The effects of physical distance and response latency on persuasion in computer-mediated communication and human–computer communication. *Journal of Experimental Psychology: Applied*, 5(4), 379–392. <https://doi.org/10.1037/1076-898X.5.4.379>
- Moreau, D., & Gamble, B. (2022). Conducting a meta-analysis in the age of open science: Tools, tips, and practical recommendations. *Psychological Methods*, 27(3), 426–432. <https://doi.org/10.1037/met0000351>
- Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>

- Mottarella, K. E., Fritzsche, B. A., & Cerabino, K. C. (2004). What do students want in advising? A Policy Capturing Study. *NACADA Journal*, 24(1 & 2), 48–61. <https://doi.org/10.12930/0271-9517-24.1-2.48>
- Mulac, A., Erlandson, K. T., Farrar, W. J., Hallett, J. S., Molloy, J. L., & Prescott, M. E. (1998). “Uh-huh. What’s that all about?” *Communication Research*, 25(6), 641–668. <https://doi.org/10.1177/009365098025006004>
- Narducci, F., Basile, P., De Gemmis, M., Lops, P., & Semeraro, G. (2020). An investigation on the user interaction modes of conversational recommender systems for the music domain. *User Modeling and User-Adapted Interaction*, 30(2), 251–284. <https://doi.org/10.1007/s11257-019-09250-7>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In B. Adelson (Ed.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). Association for Computing Machinery. <https://doi.org/10.1145/191666.191703>
- Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on authenticity in chat bots. *Multimodal Technologies and Interaction*, 2(3), 60. <https://doi.org/10.3390/mti2030060>
- Nguyen, M., Ferm, L., Quach, S., Pontes, N., & Thaichon, P. (2023). Chatbots in frontline services and customer experience: An anthropomorphism perspective. *Psychology & Marketing*, 40(11), 2201–2225. <https://doi.org/10.1002/mar.21882>
- Nudelman, G., & Otto, K. (2020). The development of a new generic risk-of-bias measure for systematic reviews of surveys. *Methodology*, 16(4), 278–298. <https://doi.org/10.5964/meth.4329>
- Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5. <https://doi.org/10.3389/frobt.2018.00114>
- OpenAI. (2023). *Documentation*. <https://platform.openai.com/docs/introduction>
- OpenAI. (2024a). *ChatGPT*. OpenAI. <https://openai.com/chatgpt/>
- OpenAI. (2024b). *Hello GPT-4o*. <https://openai.com/index/hello-gpt-4o/>
- Orsingher, C., Valentini, S., & de Angelis, M. (2010). A meta-analysis of satisfaction with complaint handling in services. *Journal of the Academy of Marketing Science*, 38(2), 169–186. <https://doi.org/10.1007/s11747-009-0155-z>

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. <https://doi.org/10.1136/bmj.n71>
- Park, G., Chung, J., & Lee, S. (2024). Human vs Machine-like representation in chatbot mental health counseling: The serial mediation of psychological distance and trust on compliance intention. *Current Psychology*, *43*, 4352–4363. <https://doi.org/10.1007/s12144-023-04653-7>
- Pelau, C., Niculescu, M., & Bojescu, I. (2021). Gender specific preferences towards anthropomorphic AI devices and robots. In R. Pamfilie, V. Dinu, L. Tăchiciu, D. Pleșea, & C. Vasiliu (Eds.), *7th BASIQ International Conference on New Trends in Sustainable Business and Consumption* (pp. 784–792). ASE. <https://doi.org/10.24818/BASIQ/2021/07/099>
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Poushter, J., Gubbala, S., & Austin, S. (2024, February 5). 8 charts on technology use around the world. *Pew Research Center*. <https://www.pewresearch.org/short-reads/2024/02/05/8-charts-on-technology-use-around-the-world/>
- Prahl, A., & van Swol, L. (2021). Out with the humans, in with the machines?: Investigating the behavioral and psychological effects of replacing human advisors with a machine. *Human-Machine Communication*, *2*, 209–234. <https://doi.org/10.30658/hmc.2.11>
- R Core Team. (2023). *R* (Version 4.3.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. <https://doi.org/10.1007/s11747-019-00710-5>
- Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, *151*, 102630. <https://doi.org/10.1016/j.ijhcs.2021.102630>

- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. CSLI Publications.
- Regulation (EU) 2024/1689 Laying down Harmonised Rules on Artificial Intelligence, Pub. L. No. 2024/1689 (2024). <http://data.europa.eu/eli/reg/2024/1689/oj>
- Reis, H. T., & Clark, M. S. (2013). Responsiveness. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 400–423). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195398694.013.0018>
- Reuters. (2024, August 29). OpenAI says ChatGPT’s weekly users have grown to 200 million. *Reuters*. <https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/>
- Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. (1995). The well-built clinical question: A key to evidence-based decisions. *ACP Journal Club*. <https://doi.org/10.7326/acpjc-1995-123-3-a12>
- Rivas, P., & Zhao, L. (2023). Marketing with ChatGPT: Navigating the ethical terrain of GPT-based chatbot technology. *AI*, 4(2), 375–384. <https://doi.org/10.3390/ai4020019>
- Robinette, P., Howard, A. M., & Wagner, A. R. (2015). Timing Is key for robot trust repair. In A. Tapus, E. André, J.-C. Martin, F. Ferland, & M. Ammi (Eds.), *Social Robotics. ICSR 2015. Lecture Notes in Computer Science* (Vol. 9388, pp. 574–583). Springer. https://doi.org/10.1007/978-3-319-25554-5_57
- Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, 26(2), 141–160. <https://doi.org/10.1037/met0000300>
- Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, 6(58), 5425. <https://doi.org/10.1126/scirobotics.abj5425>
- Roos, C. A., Postmes, T., & Koudenburg, N. (2023). Feeling heard: Operationalizing a key concept for social relations. *PLOS ONE*, 18(11), e0292865. <https://doi.org/10.1371/journal.pone.0292865>
- Rosseel, Y., Jorgensen, Terrence, D., & Rockwood, N. (2021). *Package “lavaan”: latent variable analysis* [Computer software]. <https://doi.org/10.18637/jss.v048.i02>
- Sartorius, N. (2006). The meanings of health and its promotion. *Croatian Medical Journal*, 47(4), 662–664.
- Schank, R. C., & Abelson, R. P. (1975). Scripts, plans, and knowledge. *Proceedings of the 4th International Joint Conference on Artificial Intelligence - Volume 1*, 151–157.

- Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, 32(3), 736–751. <https://doi.org/10.1287/isre.2021.1015>
- Scherer, A., Fischer, P. M., Schmitt, B., & Egli, J. (2020, May). *Wow, that’s great! The effect of phatic cues in chatbot conversations*. European Marketing Academy Conference. EMAC, Budapest, Hungary. <https://www.alexandria.unisg.ch/262527/>
- Schindler, D., Maiberger, T., Koschate-Fischer, N., & Hoyer, W. D. (2024). How speaking versus writing to conversational agents shapes consumers’ choice and choice satisfaction. *Journal of the Academy of Marketing Science*, 52(3), 634–652. <https://doi.org/10.1007/s11747-023-00987-7>
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 103. <https://doi.org/10.9781/ijimai.2017.09.001>
- Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, 37(3), 875–900. <https://doi.org/10.1080/07421222.2020.1790204>
- Schwede, M., Zierau, N., Janson, A., Hammerschmidt, M., & Leimeister, J. M. (2022, December 9). “I will follow you!” – How recommendation modality impacts processing fluency and purchase intention. *Forty-Third International Conference on Information Systems, Copenhagen, 2022*. <https://dx.doi.org/10.2139/ssrn.4239822>
- Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: A psychological approach. *Universal Access in the Information Society*, 4(2), 146–155. <https://doi.org/10.1007/s10209-005-0118-1>
- Seeger, A.-M., & Heinzl, A. (2021). Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service? *ECIS 2021 Research Papers*, 12. https://aisel.aisnet.org/ecis2021_rp/12
- Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with human-like conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22(4), 931–967. <https://doi.org/10.17705/1JAIS.00685>
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ (Clinical Research Ed.)*, 350, 7647. <https://doi.org/10.1136/bmj.g7647>

- Shanahan, M. (2024). Talking about large language models. *Communications of the ACM*, 67(2), 68–79. <https://doi.org/10.1145/3624724>
- Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, 22(1), 29–49. <https://doi.org/10.21248/jlcl.22.2007.88>
- Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, 115, 14–24. <https://doi.org/10.1016/j.jbusres.2020.04.030>
- Silva, G. R. S., & Canedo, E. D. (2024). Towards user-centric guidelines for chatbot conversational design. *International Journal of Human–Computer Interaction*, 40(2), 98–120. <https://doi.org/10.1080/10447318.2022.2118244>
- Singh, B., Olds, T., Brinsley, J., Dumuid, D., Virgara, R., Matricciani, L., Watson, A., Szeto, K., Eglitis, E., Miatke, A., Simpson, C. E. M., Vandelanotte, C., & Maher, C. (2023). Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. *Npj Digital Medicine*, 6(1), 1–10. <https://doi.org/10.1038/s41746-023-00856-1>
- Solomon, J., & Wash, R. (2014). Human-what Interaction? Understanding user source orientation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 422–426. <https://doi.org/10.1177/1541931214581088>
- Song, S. W., & Shin, M. (2024). Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of E-commerce: The moderating role of avatar familiarity. *International Journal of Human–Computer Interaction*, 40(2), 441–456. <https://doi.org/10.1080/10447318.2022.2121038>
- Spence, P. R., Westerman, D., Edwards, C., & Edwards, A. (2014). Welcoming our robot overlords: Initial expectations about interaction with a robot. *Communication Research Reports*, 31(3), 272–280. <https://doi.org/10.1080/08824096.2014.924337>
- Stein, J.-P., Messingschlager, T., & Hutmacher, F. (2023). Künstliche Intelligenz. In M. Appel, F. Hutmacher, C. Mengelkamp, J.-P. Stein, & S. Weber (Eds.), *Digital ist besser?! Psychologie der Online- und Mobilkommunikation* (pp. 247–260). Springer. https://doi.org/10.1007/978-3-662-66608-1_17
- Steinhoff, L., Arli, D., Weaven, S., & Kozlenkova, I. V. (2019). Online relationship marketing. *Journal of the Academy of Marketing Science*, 47(3), 369–393. <https://doi.org/10.1007/S11747-018-0621-6>

- Stokel-Walker, C., & Van Noorden, R. (2023). The promise and peril of generative AI. *Nature*, *614*, 214–216. <https://doi.org/10.1038/d41586-023-00340-6>
- Straßmann, C., Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2018). With or against each other? The influence of a virtual agent's (non)cooperative behavior on user's cooperation behavior in the Prisoners' Dilemma. *Advances in Human-Computer Interaction*, *2018*, 2589542. <https://doi.org/10.1155/2018/2589542>
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press. <https://doi.org/10.1162/dmal.9780262562324.073>
- Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAI). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. <https://doi.org/10.1093/jcmc/zmz026>
- Sundar, S. S., Bellur, S., Oh, J., Jia, H., & Kim, H.-S. (2016). Theoretical importance of contingency in human-computer interaction. *Communication Research*, *43*(5), 595–625. <https://doi.org/10.1177/0093650214534962>
- Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–9. <https://doi.org/10.1145/3290605.3300768>
- Sundar, S. S., & Liao, M. (2023). Calling BS on ChatGPT: Reflections on AI as a communication source. *Journalism & Communication Monographs*, *25*(2), 165–180. <https://doi.org/10.1177/15226379231167135>
- Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, *51*(1), 52–72. <https://doi.org/10.1111/j.1460-2466.2001.tb02872.x>
- Sundar, S. S., Xu, Q., & Dou, X. (2019). The Role of technology in online persuasion: A MAIN model perspective. In *Advertising Theory* (2nd ed.). Routledge.
- Tingley, D., Yamamoto, K., Keele, L., Imai, K., Trinh, M., & Wong, W. (2019). *Package 'mediation': Causal mediation analysis* [Computer software]. <https://imai.princeton.edu/projects/mechanisms.html>
- Toader, D.-C., Boca, G., Toader, R., Măcelaru, M., Toader, C., Ighian, D., & Rădulescu, A. T. (2020). The effect of social presence and chatbot errors on trust. *Sustainability*, *12*(1), 256. <https://doi.org/10.3390/SU12010256>
- Trepte, S., Reinecke, L., & Schäwel, J. (2021). *Medienpsychologie* (3., erweiterte und überarbeitete Auflage). W. Kohlhammer Verlag.

- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *LIX*(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Valério, F. A. M., Guimarães, T. G., Prates, R. O., & Candello, H. (2020). Comparing users' perception of different chatbot interaction paradigms. In C. Q. Santos (Ed.), *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems* (pp. 1–10). Association for Computing Machinery. <https://doi.org/10.1145/3424953.3426501>
- Van Berlo, Z. M. C., Meijers, M. H. C., Eelen, J., Voorveld, H. A. M., & Eisend, M. (2023). When the medium is the message: A meta-analysis of creative media advertising effects. *Journal of Advertising*, *53*(2), 278–295. <https://doi.org/10.1080/00913367.2023.2186986>
- Van den Akker, O. R., Van Assen, M. A. L. M., Bakker, M., Elsherif, M., Wong, T. K., & Wicherts, J. M. (2023). Preregistration in practice: A comparison of preregistered and non-preregistered studies in psychology. *Behavior Research Methods*, *56*, 5424–5433. <https://doi.org/10.3758/s13428-023-02277-0>
- Van der Goot, M. J. (2022). Source orientation, anthropomorphism, and social presence in human-chatbot communication: How to proceed with these concepts. *Publizistik*, *67*(4), 555–578. <https://doi.org/10.1007/s11616-022-00760-w>
- Van der Goot, M. J., & Etzrodt, K. (2023). Disentangling two fundamental paradigms in human-machine communication research: Media equation and media evocation. *Human-Machine Communication*, *6*. <https://doi.org/10.30658/hmc.6.2>
- Van der Meulen, R. (2023). *Gartner identifies six ChatGPT risks legal and compliance leaders must evaluate*. Gartner. <https://www.gartner.com/en/newsroom/press-releases/2023-05-18-gartner-identifies-six-chatgpt-risks-legal-and-compliance-must-evaluate>
- Van Pinxteren, M. M. E., Pluymaekers, M., & Lemmink, J. G. A. M. (2020). Human-like communication in conversational agents: A literature review and research agenda. *Journal of Service Management*, *31*(2), 203–225. <https://doi.org/10.1108/JOSM-06-2019-0175>
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342–365. <https://doi.org/10.1287/isre.11.4.342.11872>
- Verma, V., Sharma, D., & Sheth, J. (2016). Does relationship marketing matter in online retailing? A meta-analytic approach. *Journal of the Academy of Marketing Science*, *44*(2), 206–217. <https://doi.org/10.1007/s11747-015-0429-6>

- Vevea, J. L., Coburn, K., & Sutton, A. (2019). Publication bias. In M. Borenstein, H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd edition, pp. 383–429). Russell Sage Foundation.
- Vial, G., Cameron, A.-F., Giannelia, T., & Jiang, J. (2023). Managing artificial intelligence projects: Key insights from an AI consulting firm. *Information Systems Journal*, 33(3), 669–691. <https://doi.org/10.1111/isj.12420>
- Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift Für Psychologie / Journal of Psychology*, 215(2), 104–121. <https://doi.org/10.1027/0044-3409.215.2.104>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3). <https://doi.org/10.18637/jss.v036.i03>
- Wald, R., Heijseelaar, E., & Bosse, T. (2021). Make your own: The potential of chatbot customization for the development of user trust. *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 382–387. <https://doi.org/10.1145/3450614.3463600>
- Walfisch, T., Van Dijk, D., & Kark, R. (2013). Do you really expect me to apologize? The impact of status and gender on the effectiveness of an apology in the workplace. *Journal of Applied Social Psychology*, 43(7), 1446–1458. <https://doi.org/10.1111/jasp.12101>
- Wallace, B. C., Small, K., Brodley, C. E., Lau, J., & Trikalinos, T. A. (2012). Deploying an interactive machine learning system in an evidence-based practice center: Abstract. *Proceedings of the ACM International Health Informatics Symposium (IHI)*, 819–824. <https://doi.org/10.1145/2110363.2110464>
- Wallace, R. S. (2009). The anatomy of A.L.I.C.E. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test* (pp. 181–210). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6710-5_13
- Weiss, A., & Bartneck, C. (2015). Meta analysis of the usage of the Godspeed Questionnaire series. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 381–388. <https://doi.org/10.1109/ROMAN.2015.7333568>
- Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I believe in a thing called bot: Perceptions of the humanness of “chatbots.” *Communication Studies*, 70(3), 295–312. <https://doi.org/10.1080/10510974.2018.1557233>

- Wetzel, E., & Greiff, S. (2018). The world beyond rating scales: Why we should think more carefully about the response format in questionnaires. *European Journal of Psychological Assessment, 34*(1), 1–5. <https://doi.org/10.1027/1015-5759/a000469>
- Wilson, D. B. (2023). *Practical meta-analysis effect size calculator* (Version 2023.11.27) [Computer software]. <https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html>
- Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management, 29*(5), 907–931. <https://doi.org/10.1108/JOSM-04-2018-0119>
- Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology, 39*(6), 973–990. <https://doi.org/10.1002/ejsp.595>
- Wong, B., & Bottorff, C. (2023, May 18). *Top social media statistics and trends of 2024*. Forbes. <https://www.forbes.com/advisor/business/social-media-statistics/#source>
- World Economic Forum. (2024, January 10). *Global Risks Report 2024. 19th edition. Insight report*. World Economic Forum. <https://www.weforum.org/publications/global-risks-report-2024/digest/>
- Xu, K., Chen, M., & You, L. (2023). The hitchhiker’s guide to a credible and socially present robot: Two meta-analyses of the power of social cues in human-robot interaction. *International Journal of Social Robotics, 15*, 269–295. <https://doi.org/10.1007/s12369-022-00961-3>
- Xu, K., Liu, F., Chen, X., & Lombard, M. (2023). The Media Are Social Actors paradigm and beyond: Theory, evidence, and future research. In *Research Handbook on Artificial Intelligence and Communication* (pp. 337–353). Edward Elgar. <https://doi.org/10.4337/9781803920306.00031>
- Xu, K., & Shi, J. (2024). Visioning a two-level human–machine communication framework: Initiating conversations between explainable AI and communication. *Communication Theory*, qtae016. <https://doi.org/10.1093/ct/qtae016>
- Yu, S., Xiong, J. (Jill), & Shen, H. (2024). The rise of chatbots: The effect of using chatbot agents on consumers’ responses to request rejection. *Journal of Consumer Psychology, 34*(1), 35–48. <https://doi.org/10.1002/jcpy.1330>
- Zemčik, T. (2019). A brief history of chatbots. *DEStech Transactions on Computer Science and Engineering*. <http://dx.doi.org/10.12783/dtcse/aicae2019/31439>

- Zhang, T., Feng, C., Chen, H., & Xian, J. (2022). Calming the customers by AI: Investigating the role of chatbot acting-cute strategies in soothing negative customer emotions. *Electronic Markets*, 32(4), 2277–2292. <https://doi.org/10.1007/s12525-022-00596-2>
- Zheng, T., Duan, X., Zhang, K., Yang, X., & Jiang, Y. (2023). How chatbots' anthropomorphism affects user satisfaction: The mediating role of perceived warmth and competence. In Y. Tu & M. Chi (Eds.), *E-Business. Digital Empowerment for an Intelligent Future. WHICEB 2023* (Vol. 481, pp. 96–107). Springer. https://doi.org/10.1007/978-3-031-32302-7_9
- Zierau, N., Elshan, E., Visini, C., & Janson, A. (2020). A review of the empirical literature on conversational agents and future research directions. *ICIS 2020 Proceedings*, 5. https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/5
- Zierau, N., Hildebrand, C., Bergner, A., Busquet, F., Schmitt, A., & Marco Leimeister, J. (2023). Voice bots on the frontline: Voice-based interfaces enhance flow-like consumer experiences & boost service outcomes. *Journal of the Academy of Marketing Science*, 51(4), 823–842. <https://doi.org/10.1007/s11747-022-00868-5>

Appendix

A Appendix Chapter 2¹¹

Supplementary materials, including videos of the prerecorded chatbot interactions, supporting data, and the corresponding R code, are freely accessible at <https://osf.io/w8dzv/>. The preregistrations are freely accessible at <https://aspredicted.org/y8se6.pdf> (Study 1), <https://aspredicted.org/xz5ta.pdf> (Study 2), <https://aspredicted.org/e4u3j.pdf> (amendment Study 2), and <https://aspredicted.org/cj2ec.pdf> (Study 3).

Power Analyses

The sample size of Study 1 ($N = 256$) was determined by an a priori power analysis ($f = .18$, $\alpha = .05$, $1 - \beta = .80$) for an interaction effect resulting from an ANOVA. The power analysis was based on the small interaction effect ($\eta^2_{\text{partial}} = .03$) found by Go & Sundar (2019) in an experiment on the effects of identity and conversational cues on attitude toward the website where the chatbot is placed, using real interactions with chat agents. As we expected effects in Studies 2 and 3 to be smaller than Go and Sundar's interaction effect, we determined a sample size of $N = 403$ a priori for a significant interaction effect of $f = .14$ ($\alpha = .05$, $1 - \beta = .80$).

Deviations from Preregistrations

Studies 1–3: We clarified and harmonized the wording and the order of our hypotheses and research questions in the following way:

H1: Human identity cues have a positive effect on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (g) intention to use, and (h) satisfaction ($\mu_{\text{human identity cues}} > \mu_{\text{chatbot identity cues}}$).

H2: Responsive verbal cues have a positive effect on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (g) intention to use, and (h) satisfaction ($\mu_{\text{responsive verbal cues}} > \mu_{\text{non-responsive verbal cues}}$).

H3 (Studies 1, 2): The effects of agent type on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (g) intention to use, and (h) satisfaction are mediated by social presence.

H4: The effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (g) intention to use, and (h) satisfaction are mediated by social presence.

¹¹ Most of Appendix A is included in the appendix of the published manuscript (<https://doi.org/10.30658/hmc.8.4>) or is available as supplementary material on osf (<https://osf.io/w8dzv/>).

H5: The effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (g) intention to use, and (h) satisfaction are mediated by perceived dialogue.

RQ1: Is there an interaction effect between agent type (human vs. chatbot identity cues) and responsiveness (responsive vs. no responsive verbal cues) on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (g) intention to use, and (h) satisfaction?

RQ2: Does feeling heard mediate the effects of responsiveness on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (g) intention to use, and (h) satisfaction?

Instead of computing a *t*-test and a two-way ANOVA for each outcome variable, we computed one two-way MANOVA to test H1 and H2 and to answer RQ1 at the same time. This way, we condensed the analyses while also being able to look at the two-way interaction.

Studies 1 and 2: We preregistered to compute exploratory factor analyses to assess internal consistency but instead computed Cronbach's alpha coefficients (see below).

Study 2: We had preregistered the additional exploratory research question, "Do the effects hypothesized in H1 and H2 persist when controlling for the quality of previous chatbot experience?" The results are not reported in the main text as they are not vital for answering the overall research question. The results are provided below.

Study 3: Study 3 was originally designed to additionally explore the impact of agent response time on users' responses, so response time (immediate or dynamically delayed) was included as a third experimental factor. Two exploratory research questions were formulated: To what extent does agent response time (instant vs. dynamically delayed) influence (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) intention to use, and (h) satisfaction? Are there interaction effects between agent type, responsiveness, and response time on (a) attitude, (b) likability, (c) perceived intelligence, (d) warmth, (e) competence, (f) perceived enjoyment, (g) intention to use and (h) satisfaction? Because the results were not vital to answering our overarching research question, they are provided below. In addition, we did not compute mediation models because we did not find significant effects of responsiveness on any outcome variable.

Pretest Study 3

In Studies 1 and 2, a significant proportion of participants did not recognize the human agent. Thus, prior to Study 3, we conducted a pretest with $N = 100$ participants recruited via Clickworker in August 2021 ($n_{\text{female}} = 43$, age: $M = 40.86$, $SD = 12.64$, range = 18–67). The

results indicated that fewer participants failed the agent manipulation check when we made the agent type manipulation more explicit by telling participants what agent they would see in the conversation before they watched the video. We thus adopted this approach for Study 3.

Procedure

Table A.1 Procedure Across Studies

Study 1	Study 2	Study 3
		Screening question for high school degree
	Introduction, informed consent	
	Instructions, video depending on experimental condition	
Manipulation check 1	Manipulation checks	
Questions on chatbot use (chatbot conditions)		
Technology commitment		
	Social presence	
	Feeling heard, perceived dialogue	
	Perceived moral agency ^a	
		Perceived relational commitment ^a
	Perceived intelligence, likability	
	Perceived warmth, perceived competence	
Manipulation check 2		
	Satisfaction, perceived enjoyment	
	Attitude, intention to use	
	Chatbot use (all conditions)	
	Demographic information	
	Comment (optional), debriefing, final consent	
	Separate questionnaire part ^a	
	Raffle	
	End	

Note. Core measures in bold. ^a Assessed for unrelated research projects. The separate questionnaire part included questions on attitudes towards AI and algorithms, video-conferencing use and behavior, smart speaker use.

Measures

English translations in italics. Variables were measured on 7-point Likert-type rating scales (1 = do not agree at all, 7 = fully agree), except likability, perceived intelligence (7-point semantic differentials), and categorical variables.

Attitude, adapted from Diers (2020) (Cronbach's α in Studies 1/2/3: .95/.95/.91)

1. Auf diese Weise mit einer Organisation zu kommunizieren finde ich attraktiv. *I find it attractive to communicate with an organization in this way.*
2. Ich erachte es als sinnvoll, auf diese Weise mit einer Organisation zu kommunizieren. *I find it useful to communicate with an organization in this way.*
3. Diese Art der Kommunikation mit einer Organisation finde ich interessant. *I find this way of communicating with an organization interesting.*

4. Es erscheint mir hilfreich, auf diese Weise mit einer Organisation zu kommunizieren. *I find it helpful to communicate in this way with an organization.*
5. Diese Weise mit einer Organisation zu kommunizieren hilft mir, meine Bedürfnisse zu erfüllen. *Communicating in this way with an organization helps me to meet my needs.*

Likability, adapted from Bartneck et al. (2009) (.92/.87/.90)

Bitte beurteilen Sie Ihren Eindruck von Sophie auf diesen Skalen: abgeneigt–zugeneigt, unfreundlich–freundlich, unhöflich–höflich, unangenehm–angenehm, furchtbar–nett. *Please rate your impression of Sophie on these scales: dislike–like, unfriendly–friendly, unkind–kind, unpleasant–pleasant, awful–nice.*

Perceived intelligence, adapted from Bartneck et al. (2009) (.88/.90/.92)

Bitte beurteilen Sie Ihren Eindruck von Sophie auf diesen Skalen: inkompetent–kompetent, ungebildet–unterrichtet, verantwortungslos–verantwortungsbewusst, unintelligent–intelligent, unvernünftig–vernünftig. *Please rate your impression of Sophie on these scales: incompetent–competent, ignorant–knowledgeable, irresponsible–responsible, unintelligent–intelligent, foolish–sensible.*

Perceived warmth, translated and adapted from Fiske (2018) (.87/.88/.91)

Wie [warm, vertrauenswürdig, freundlich, ehrlich, sympathisch, aufrichtig] war Sophie Ihrer Ansicht nach? *How [warm, trustworthy, friendly, honest, likable, sincere] do you think Sophie was?*

Perceived competence, translated and adapted from Fiske (2018) (.88/.86/.91)

Wie [kompetent, intelligent, fähig, effizient, durchsetzungsfähig, selbstsicher] war Sophie Ihrer Ansicht nach? *How [competent, intelligent, skilled, efficient, assertive, confident] do you think Sophie was?*

Perceived enjoyment, adapted from Diers (2020) (.90/.89/.91)

1. Das Gespräch ruft bei mir positive Gefühle hervor. *The conversation evokes positive feelings in me.*
2. Ich fand das Gespräch unterhaltsam. *I found the conversation entertaining.*
3. Es hat mir Spaß gemacht, das Gespräch zu verfolgen. *I enjoyed reading the conversation.*

Satisfaction, translated and adapted from Lagace et al. (1991) (.92/.91/.94)

1. Ich wäre zufrieden mit Sophies Empfehlungen für Studiengänge. *I would be happy with Sophie's recommendations for courses of study.*
2. Ich wäre zufrieden mit der Art und Weise, wie Sophie mit Marc gesprochen hat. *I would be satisfied with the way Sophie spoke to Marc.*

3. Ich wäre zufrieden mit den Informationen, die Sophie Marc gegeben hat. *I would be satisfied with the information Sophie gave Marc.*
4. Das Gespräch mit Sophie hätte mich zufrieden gestellt. *I would be satisfied with the conversation with Sophie.*

Intention to use, adapted from Diers (2020) (.96/.96/.97)

1. Wenn eine Organisation diese Möglichkeit der Kommunikation anbietet, werde ich diese auch nutzen. *If an organization offers this possibility of communication, I will use it.*
2. Wenn ich die Möglichkeit erhalte, auf diese Weise mit einer Organisation zu kommunizieren, werde ich das auch tun. *If I have the opportunity to communicate with an organization in this way, I will.*
3. Diese Weise, mit einer Organisation zu kommunizieren, werde ich mit sehr hoher Wahrscheinlichkeit nutzen. *I am very likely to use this way of communicating with an organization.*
4. Sobald sich diese Weise, mit einer Organisation zu kommunizieren, etabliert, werde ich diese bevorzugt nutzen. *Once this way of communicating with an organization is established, it will be my preferred method.*

Social presence, translated and adapted from Gefen & Straub (2004) (.94/.94/.96)

Im Gespräch bestand ein Gefühl [des menschlichen Kontakts, der Persönlichkeit, der Geselligkeit, menschlicher Wärme, menschlicher Empfindsamkeit]. *There was a sense of [human contact, personalness, sociability, human warmth, human sensitivity] in the interaction.*

Perceived dialogue, translated and adapted from Sundar et al. (2016) (.85/.82/.86)

1. Ich hatte das Gefühl, dass Sophie in einem aktiven Dialog mit Marc stand. *I had the feeling that Sophie was in an active dialogue with Marc.*
2. Die Interaktionen von Marc mit Sophie fühlten sich wie ein Hin- und Her-Gespräch an. *I felt that Sophie and Marc were involved in a joint task when choosing a program*
3. Ich hatte das Gefühl, dass Sophie und Marc bei der Auswahl eines Studiengangs in eine gemeinsame Aufgabe eingebunden waren. *Marc's interactions with Sophie felt like a back-and-forth conversation.*
4. Sophie reagierte schnell auf Marcs Angaben und Wünsche. *Sophie was quick to respond to Marc's input and requests.*
5. Ich hatte das Gefühl, dass Sophie Marcs individuelle Wünsche berücksichtigte. *I felt that Sophie took Marc's individual wishes into account.*

Feeling heard, translated and adapted from Roos et al. (2023) (.84/.82/.86)

1. Marc fühlte sich gehört. *Marc felt heard.*
2. Marc konnte sagen, was er wirklich sagen wollte. *Marc was able to say what he really wanted to say.*
3. Sophie schien sich mehr um etwas anderes zu kümmern als um das, was Marc gesagt hat. *Sophie seemed to care more about something else than what Marc said.*
4. Sophie hörte Marc zu. *Sophie listened to Marc.*
5. Sophie versuchte, sich in Marcs Lage zu versetzen. *Sophie tried to put herself in Marc's shoes.*
6. Sophie wirkte unsensibel gegenüber Marcs Gedanken und Gefühlen. *Sophie seemed insensitive to Marc's thoughts and feelings.*
7. Sophie behandelte Marc mit Respekt. *Sophie treated Marc with respect.*
8. Sophie und Marc verstanden sich gegenseitig. *Sophie and Marc understood each other.*

Agent type manipulation check, translated and adapted from Go & Sundar (2019)

Wenn Sie an den Chatverlauf zurückdenken, den Sie gerade gesehen haben: Mit wem hat Marc gesprochen? *If you think back to the chat interaction you just saw: Who was Marc talking to?* [1 = der Studienberaterin *the study advisor* Sophie, 2 = der Professorin *the professor* Sophie, 3 = dem Chatbot *the chatbot* Sophie, 4 = der Ärztin *the doctor* Sophie, 5 = weiß nicht *don't know*]

Responsiveness manipulation check, translated and adapted from De Ruyter & Wetzels (2000) (.77/.65/.59)

1. Study 1: Sophie benutzte bestätigende Ausdrücke, um anzuzeigen, dass sie Marc wirklich zuhörte. *Sophie used affirmative expressions to indicate that she was really listening to Marc.*
Studies 2, 3: Sophie benutzte bestätigende Ausdrücke, um anzuzeigen, dass sie Marc zuhörte. *Sophie used affirmative expressions to indicate that she was listening to Marc.*
2. Study 1: Sophie hat das was Marc gesagt hat, in ihrer Antwort angemessen aufgegriffen. *Sophie appropriately picked up on what Marc said in her response.*
Studies 2, 3: Sophie hat das was Marc gesagt hat, in ihrer Antwort aufgegriffen. *Sophie picked up on what Marc said in her response.*

Chatbot use frequency

Study 1: Wie oft haben Sie Chatbots in den letzten 12 Monaten durchschnittlich pro Monat verwendet? [gar nicht, einmal pro Monat, zwei- bis dreimal pro Monat, mehr als

dreimal pro Monat] *How often have you used chatbots on average per month in the last 12 months?* [not at all, once a month, two to three times a month, more than three times a month]

Studies 2, 3: Wie oft haben Sie in den letzten drei Monaten einen Chatbot verwendet? [gar nicht, einmal, zwei- bis dreimal, mehr als dreimal] *How often have you used a chatbot in the last three months?* [Not at all, once, two to three times, more than three times] add

Previous chatbot experience

Denken Sie bitte an die Erfahrungen, die Sie in den letzten drei Monaten mit Chatbots gemacht haben. Wie würden Sie Ihre Erfahrungen im Allgemeinen bewerten? [1 = überhaupt nicht gut, 7 = sehr gut] *Please think about the experiences you have had with chatbots in the last three months. How would you rate your experience in general?* [1 = not good at all, 7 = very good]

Stimulus Materials

Original Chat Conversations

Not Responsive Human Condition

Beraterin: Hallo, ich bin Sophie, eine Mitarbeiterin der Studienberatung an der EDU-Hochschule. Ich beantworte Fragen rund um unser Studienangebot. Außerdem kann ich dir helfen, den idealen Studiengang für dich zu finden. Unsere Hochschule bietet berufsbegleitende Bachelor- und Masterstudiengänge in den Bereichen Wirtschaft und Management an. Mit wem spreche ich?

Nutzer: Ich heiße Marc

Beraterin: Schön, dass du dich für ein Studium bei uns interessierst.

Beraterin: Hast du schon einen speziellen Studiengang im Blick oder möchtest du, dass ich dir mit der Auswahl eines Studiengangs helfe?

Nutzer: Ich hätte gerne eine Beratung.

Beraterin: Ich werde dir nun einige Fragen stellen und dir auf der Grundlage deiner Antworten Vorschläge für einen Studiengang machen. Wir fangen mit Fragen zu deinem höchsten Bildungsabschluss an. Welches ist dein höchster Bildungsabschluss?

Nutzer: Mein höchster Abschluss ist ein Bachelor of Science.

Beraterin: In welchem Fach hast du deinen höchsten Bildungsabschluss erlangt?

Nutzer: in Betriebswirtschaftslehre

Beraterin: Hast du dich in deinem Studium auf bestimmte Fächer spezialisiert?

Nutzer: Ja, auf Marketing und Rechnungswesen

Beraterin: Danke für deine Angaben. Welchen Abschluss möchtest du an der EDU-Hochschule erlangen?,

Nutzer: Ich möchte gerne einen berufsbegleitenden Master machen.

Beraterin: Jetzt stelle ich dir noch Fragen zu deinem aktuellen Arbeitsverhältnis. Neben einem wirtschaftswissenschaftlich ausgerichteten Bachelorabschluss ist eine Berufstätigkeit Voraussetzung, um bei uns studieren zu können. Bist du berufstätig?

Nutzer: Ja, ich arbeite Vollzeit in einem mittelständischen Unternehmen.

Beraterin: Und in welchem Bereich arbeitest du momentan?

Nutzer: Im Marketing

Beraterin: Zu guter Letzt benötige ich noch Informationen zu deinen fachlichen Kenntnissen. Nach deiner eigenen Einschätzung, wo möchtest du dich fachlich noch weiterentwickeln?

Nutzer: Hm, ich möchte auf jeden Fall mein Wissen im Bereich Kunden- und Marktanalysen erweitern.

Beraterin: Danke für deine Antworten. Die folgenden Studiengänge kommen für dich in Frage: Master of Science (M.Sc.) Marketing & Communication, Master of Science (M.Sc.) Sales Management. Auf unserer Internetseite kannst du sie dir unter Wirtschaft & Management/Master-Studiengänge genauer anschauen. Wende dich bei weiteren Fragen gerne an meine Kolleginnen und Kollegen. Deren Kontaktdaten findest du ebenfalls auf der Internetseite unter Kontakt.

Nutzer: Vielen Dank für die Vorschläge! Dann sehe ich mir die Studiengänge mal an.

Beraterin: Danke für dein Interesse an einem Studiengang an der EDU-Hochschule.

Responsive Chatbot Condition

Beraterin: Hallo, ich bin Sophie, der Chatbot der Studienberatung an der EDU-Hochschule. Ich beantworte Fragen rund um unser Studienangebot. Außerdem kann ich dir helfen, den idealen Studiengang für dich zu finden. Unsere Hochschule bietet berufsbegleitende Bachelor- und Masterstudiengänge in den Bereichen Wirtschaft und Management an. Mit wem spreche ich?

Nutzer: Ich heiße Marc

Beraterin: Schön, dass du dich für ein Studium bei uns interessierst, Marc. Hast du schon einen speziellen Studiengang im Blick oder möchtest du, dass ich dir mit der Auswahl eines Studiengangs helfe?

Nutzer: Ich hätte gerne eine Beratung.

Beraterin: Verstanden. Ich werde dir nun einige Fragen stellen und dir auf der Grundlage deiner Antworten Vorschläge für einen Studiengang machen. Wir fangen mit Fragen zu deinem höchsten Bildungsabschluss an.,

Beraterin: Welches ist dein höchster Bildungsabschluss?

Nutzer: Mein höchster Abschluss ist ein Bachelor of Science.

Beraterin: Alles klar. In welchem Fach hast du deinen Bachelor of Science erlangt?

Nutzer: in Betriebswirtschaftslehre

Beraterin: Hast du dich in deinem Bachelorstudium der Betriebswirtschaftslehre auf bestimmte Fächer spezialisiert? Wenn ja, auf welche?

Nutzer: Ja, auf Marketing und Rechnungswesen

Beraterin: Danke für deine Angaben. Welchen Abschluss möchtest du an der EDU-Hochschule erlangen?

Nutzer: Ich möchte gerne einen berufsbegleitenden Master machen.

Beraterin: Ist notiert. Jetzt stelle ich dir noch Fragen zu deinem aktuellen Arbeitsverhältnis. Neben einem wirtschaftswissenschaftlich ausgerichteten Bachelorabschluss ist eine Berufstätigkeit Voraussetzung, um bei uns studieren zu können. Bist du berufstätig?

Nutzer: Ja, ich arbeite Vollzeit in einem mittelständischen Unternehmen.

Beraterin: Mhm. Und in welchem Bereich dieses mittelständischen Unternehmens arbeitest du momentan in Vollzeit?

Nutzer: Im Marketing

Beraterin: Verstehe. Zu guter Letzt benötige ich noch Informationen zu deinen fachlichen Kenntnissen. Nach deiner eigenen Einschätzung, wo möchtest du dich fachlich noch weiterentwickeln?

Nutzer: Hm, ich möchte auf jeden Fall mein Wissen im Bereich Kunden- und Marktanalysen erweitern.,

Beraterin: Danke für deine Antworten. Hmm, lass mich kurz nachdenken. Aufgrund deiner Angaben kommen diese beiden Studiengänge für dich in Frage: Master of Science (M.Sc.) Marketing & Communication, Master of Science (M.Sc.) Sales Management. Auf unserer Internetseite kannst du sie dir unter Wirtschaft & Management/Master-Studiengänge genauer anschauen. Wende dich bei weiteren Fragen gerne an meine Kolleginnen und Kollegen. Deren Kontaktdaten findest du ebenfalls auf der Internetseite unter Kontakt.

Nutzer: Vielen Dank für die Vorschläge! Dann sehe ich mir die Studiengänge mal an.

Beraterin: Viel Spaß dabei! Danke für dein Interesse an einem Studiengang an der EDU-Hochschule. Ich hoffe, dass ich dir weiterhelfen konnte.

Translated Chat Conversations

Responsive Chatbot Condition

Advisor: Hello, I am Sophie, the chatbot of the student advisory service at EDU University. I answer questions about our study programs. I can also help you find the ideal degree program for you. Our university offers part-time bachelor's and master's degree programs in business and management. Who am I talking to?

User: My name is Marc

Advisor: We're glad you're interested in studying with us, Marc. Do you already have a specific major in mind, or would you like me to help you select a major?

User: I would like to have a consultation.

Advisor: Understood. I'm now going to ask you some questions and make suggestions for a course of study based on your answers. We'll start with questions about your highest level of education. What is your highest educational qualification?

User: My highest degree is a Bachelor of Science.

Advisor: All right. In which subject did you obtain your Bachelor of Science?

User: in business administration.

Advisor: Did you specialize in certain subjects during your bachelor's degree in business administration? If so, which ones?

User: Yes, in marketing and accounting

Advisor: Thank you for your information. What degree do you want to get at EDU University?

User: I would like to work part-time on a master's degree.

Advisor: Noted. Now, I'll ask you some questions about your current employment. In addition to a bachelor's degree in economics, a job is a prerequisite for studying with us. Are you employed?

User: Yes, I work full-time in a medium-sized company.

Advisor: Mhm. And in which area of this medium-sized company are you currently working full-time?

User: In Marketing

Advisor: I see. Finally, I need some information about your professional skills. According to your advisor assessment, where would you like to develop further professionally?

User: Hm, I definitely want to expand my knowledge in customer and market analysis.

Advisor: Thanks for your answers.

Hmm, let me think for a moment. Based on the information you have provided, both of these programs are suitable for you: Master of Science (M.Sc.) Marketing & Communication, Master of Science (M.Sc.) Sales Management. You can take a closer look at them on our website under Business & Management/Master's programs. If you have

any further questions, please do not hesitate to contact my colleagues. You can also find their contact details on the website under Contact.

User: Thanks for the suggestions! Then I'll take a look at the study programs.

Advisor: Have fun with it! Thank you for your interest in a degree program at EDU University. I hope that I could help you further.

Not Responsive Human Condition

Advisor: Hello, I'm Sophie, a member of the student advising staff at EDU University. I answer questions about our study programs. I can also help you find the ideal degree program for you. Our university offers part-time bachelor's and master's degree programs in business and management. Who am I talking to?

User: My name is Marc

Advisor: We're glad you're interested in studying with us. Do you already have a specific major in mind, or would you like me to help you select a major?

User: I would like to have a consultation.

Advisor: I'm now going to ask you some questions and make suggestions for a course of study based on your answers. We'll start with questions about your highest level of education. What is your highest educational qualification?

User: My highest degree is a Bachelor of Science.

Advisor: In which subject did you obtain your highest educational qualification?

User: in business administration

Advisor: Did you specialize in certain subjects during your studies?

User: Yes, in marketing and accounting

Advisor: Thank you for your information. What degree do you want to get at EDU University?

User: I would like to work part-time on a master's degree.

Advisor: Now, I'll ask you some questions about your current employment. In addition to a bachelor's degree in economics, you must be employed to study with us. Are you employed?

User: Yes, I work full-time in a medium-sized company.

Advisor: And what area are you working in at the moment?

User: In marketing

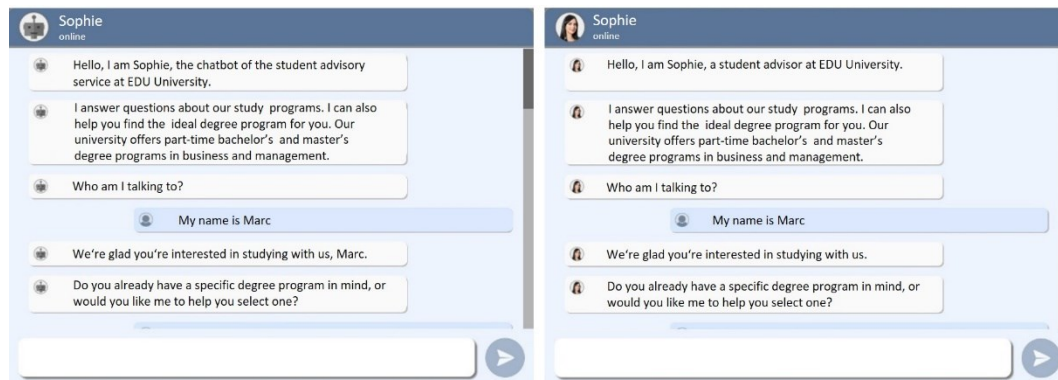
Advisor: Finally, I need information about your professional knowledge. According to your advisor assessment, where would you like to develop further professionally?

User: Hm, I definitely want to expand my knowledge in customer and market analysis.

Advisor: Thanks for your answers. The following courses of study are suitable for you: Master of Science (M.Sc.) Marketing & Communication, Master of Science (M.Sc.) Sales Management. You can take a closer look at them on our website under Business & Management/Master's programs. If you have any further questions, please do not hesitate to contact my colleagues. You can also find their contact details on the website under Contact.

User: Thanks for the suggestions! Then I'll take a look at the study programs.

Advisor: Thank you for your interest in a degree program at EDU University.

Figure A.1 Screenshots of Exemplary Interactions in Chatbot and Human Conditions

Additional Analyses

Across Studies

Table A.2 Means and Standard Deviations for Variables Across Studies

Variable	Study 1		Study 2		Study 3		Pooled Data	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Dependent Variables								
Attitude	4.79	1.58	4.18	1.67	4.76	1.61	4.53	1.65
Likability	5.68	1.08	5.53	1.06	5.72	0.96	5.63	1.04
Perceived intelligence	5.90	0.88	5.43	1.13	5.66	1.00	5.63	1.04
Warmth	5.35	0.99	5.01	1.11	5.29	1.11	5.19	1.09
Competence	5.73	0.90	5.43	0.98	5.54	1.06	5.54	1.00
Perceived enjoyment	4.29	1.42	3.78	1.51	4.55	1.50	4.18	1.52
Satisfaction	5.34	1.38	4.83	1.52	5.30	1.39	5.12	1.46
Intention to use	4.74	1.76	4.04	1.84	4.80	1.67	4.49	1.80
Mediators								
Social presence	3.80	1.52	3.58	1.51	4.35	1.52	3.91	1.55
Perceived dialogue	5.28	1.16	5.13	1.19	5.49	1.10	5.29	1.16
Feeling heard	5.43	0.92	5.30	0.95	5.48	0.97	5.39	0.95

Note. $N_1 = 253$, $N_2 = 401$, $N_3 = 351$, $N_{\text{total}} = 1,005$.

Table A.3 Bivariate Correlations Between Variables Across Studies

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Attitude	—										
2. Likability	.39	—									
	.43										
	.49										
3. Perceived intelligence	.47	.70	—								
	.38	.59									
	.54	.73									
4. Warmth	.57	.69	.63	—							
	.56	.68	.52								
	.60	.74	.71								
5. Competence	.48	.44	.67	.66	—						
	.50	.49	.54	.67							
	.59	.64	.76	.80							
6. Perceived enjoyment	.61	.57	.49	.65	.45	—					
	.57	.46	.31	.61	.49						
	.65	.59	.56	.73	.60						
7. Satisfaction	.61	.58	.66	.72	.68	.61	—				
	.65	.58	.52	.72	.68	.63					
	.70	.63	.64	.76	.73	.71					
8. Intention to use	.89	.36	.41	.48	.44	.56	.53	—			
	.86	.38	.32	.49	.42	.54	.57				
	.91	.48	.50	.57	.54	.59	.63				
9. Social presence	.54	.51	.48	.68	.50	.71	.49	.57	—		
	.51	.53	.42	.63	.50	.66	.46	.61			
	.59	.52	.50	.67	.55	.70	.57	.59			
10. Perceived dialogue	.52	.56	.51	.73	.61	.55	.50	.71	.57	—	
	.49	.56	.46	.69	.59	.54	.43	.64	.63		
	.51	.61	.64	.72	.73	.59	.48	.68	.57		
11. Feeling heard	.52	.56	.53	.71	.59	.46	.45	.67	.53	.74	—
	.43	.55	.42	.67	.58	.47	.36	.59	.41	.79	
	.46	.66	.62	.74	.72	.53	.42	.66	.51	.79	

Note. Pearson's correlations r . Grey shaded cells: 1st line = Study 1 ($N = 253$), 2nd line = Study 2 ($N = 401$), 3rd line = Study 3 ($N = 351$). All correlations are significant at $p < .001$.

Study 2

Due to the conclusions of the first study, we preregistered the additional research question, “Do the effects hypothesized in H1 and H2 persist when controlling for the quality of previous chatbot experience?”. Multivariate linear regression models among the 170 participants who had used a chatbot before, with participants’ perceived experience quality as a control variable, yielded positive agent type effects on all outcomes except satisfaction, regardless of the quality of previous chatbot experiences (Table A.4).

Table A.4 Study 2: Regression coefficients of Agent Type on Outcomes of Interest

Variable	Model 1		Model 2	
	$B_{agent\ type}$	$SE_{agent\ type}$	$B_{agent\ type}$ $B_{quality}$	$SE_{agent\ type}$ $SE_{quality}$
Attitude	0.68**	.24	0.55**	.19
R^2	.04		.06	.06
Likability	0.27°	.15	0.22	.15
R^2	.01		.11	.05
Perceived intelligence	0.40*	.17	0.38*	.17
R^2	.03		.05	.05
Warmth	0.41**	.15	0.36*	.14
R^2	.04		.14	.04
Competence	0.37**	.14	0.33*	.14
R^2	.03		.12	.04
Perceived enjoyment	0.02	.22	−.08	.20
R^2	.01		.20	.06
Intention to use	0.96***	.27	0.81***	.22
R^2	.06		.39	.07
Satisfaction	0.27	.21	0.19	.20
R^2	.00		.15	.06

Note. $N = 170$. Model 1 includes the predictor agent type (0 = chatbot, 1 = human). Model 2 additionally includes the predictor chatbot experience quality (1 = not good at all to 7 = very good).

° $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

We did not find significant responsiveness effects, regardless of participants’ experience quality (Table A.5). Additional exploratory two-way ANOVAs showed that having used a chatbot before significantly moderated the agent type effects on the outcome variables (Table A.6). Exploratory posthoc analyses signaled that participants in the human conditions indicated higher values on all outcome variables except perceived intelligence, competence, and

perceived enjoyment when they had used (vs. not used) a chatbot before. We did not find significant effects of previous chatbot use among participants in the chatbot conditions.

Table A.5 Study 2: Regression Coefficients of Responsiveness on Outcomes of Interest

Variable	Model 1		Model 2	
	$B_{responsiveness}$	$SE_{responsiveness}$	$B_{responsiveness}$ $B_{quality}$	$SE_{responsiveness}$ $SE_{quality}$
Attitude	0.03	.24	0.20 0.61***	.19 .06
R^2	.01		.37	
Likability	-0.03	.15	0.03 0.20***	.15 .05
R^2	.01		.09	
Perceived intelligence	-0.03	.17	0.01 0.12*	.17 .05
R^2	.01		.02	
Warmth	0.00	.15	0.06 0.22***	.14 .05
R^2	.01		.11	
Competence	0.05	.14	0.10 0.19***	.14 .04
R^2	.01		.10	
Perceived enjoyment	-0.08	.22	0.04 0.41***	.20 .06
R^2	.01		.20	
Intention to use	-0.02	.28	0.17 0.68***	.23 .07
R^2	.01		.34	
Satisfaction	0.27	.21	0.37° 0.35***	.20 .06
R^2	.00		.16	

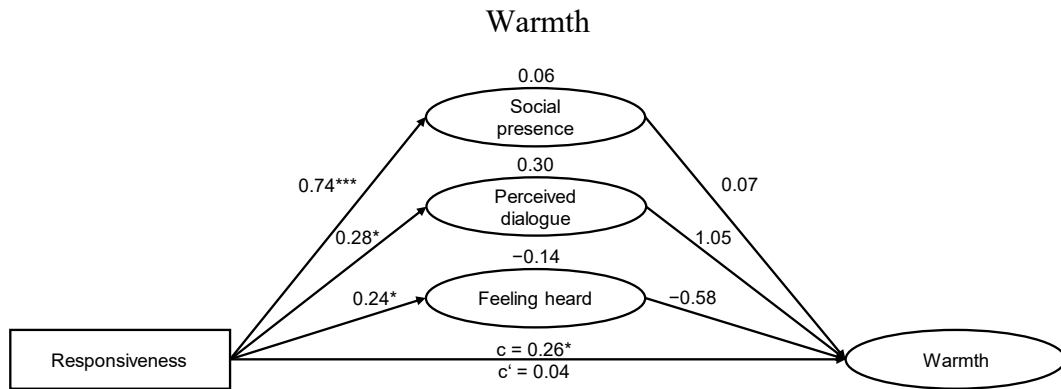
Note. $N = 170$. Model 1 includes the predictor responsiveness (0 = low, 1 = high). Model 2 additionally includes the predictor chatbot experience quality (1 = not good at all to 7 = very good).

° $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

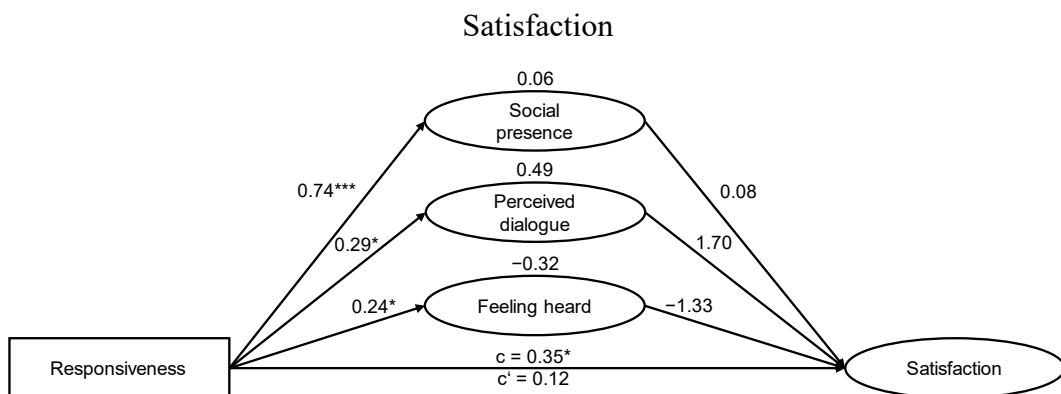
Table A.6 Study 2: Exploratory Results from Two-way ANOVA for Outcomes of Interest

Variable	Chatbot		Human		F	p	η^2
	No	Yes	No	Yes			
Attitude	4.11 _a	4.11 _a	3.89 _a	4.80 _b	7.13	.008	0.02
Likability	5.75 _{ab}	5.48 _{ab}	5.17 _a	5.75 _b	15.76	< .001	0.04
Perceived intelligence	5.54 _a	5.24 _a	5.32 _a	4.64 _a	7.26	.007	0.02
Warmth	5.22 _a	4.88 _{abc}	4.66 _b	5.29 _c	19.50	< .001	0.05
Competence	5.56 _a	5.26 _a	5.31 _a	5.63 _a	9.63	.002	0.02
Perceived enjoyment	3.92 _a	3.87 _a	3.45 _a	3.89 _a	2.53	.113	0.01
Intention to use	3.89 _a	3.92 _a	3.75 _a	4.88 _b	9.03	.003	0.02
Satisfaction	5.05 _a	4.80 _{ab}	4.43 _b	5.06 _a	8.42	0.004	0.02

Note. $N = 401$. Agent type: 0 = chatbot, 1 = human. Previous chatbot use: 0 = no, 1 = yes. Means with different subscripts within rows differ with $p < .05$.

Figure A.2 Study 2: Parallel Multiple Mediator Model for the Effect of Responsiveness on

Note. $df = 265$. $\chi^2 = 1144.720$, $p < .001$, CFI = .858, RMSEA = .091, $CI_{RMSEA}(.086, .097)$, SRMR = .076. The coefficients presented are unstandardized linear regression coefficients. $R^2_{Warmth} = .702$, $R^2_{Social\ presence} = .067$, $R^2_{Perceived\ dialogue} = .012$, $R^2_{Feeling\ heard} = .013$
 *** $p < .001$, ** $p < .01$, * $p < .05$.

Figure A.3 Study 2: Parallel Multiple Mediator Model for the Effect of Responsiveness on

Note. $df = 220$. $\chi^2 = 762.096$, $p < .001$, CFI = .906, RMSEA = .079, $CI_{RMSEA}(.073, .085)$, SRMR = .071. The coefficients presented are unstandardized linear regression coefficients. $R^2_{Satisfaction} = .600$, $R^2_{Social\ presence} = .066$, $R^2_{Perceived\ dialogue} = .012$, $R^2_{Feeling\ heard} = .013$.
 *** $p < .001$, ** $p < .01$, * $p < .05$.

B Appendix Chapter 3¹²

Supplementary materials, including videos of exemplary chatbot interactions, supporting data, and the corresponding R code, are freely accessible at <https://osf.io/j8hde/>. The preregistration is freely accessible at <https://osf.io/shb6z>.

Pretest

Procedure

1. Introduction, informed consent
2. Scenario, instructions, attention check 1
3. Chatbot interaction depending on condition
4. **Open-ended feedback, feedback on technical issues**
5. Manipulation check, attention check 2
6. **User experience questionnaires: User Experience Questionnaire (UEQ), System Usability Scale (SUS), Chatbot Usability Questionnaire (CUQ)**, affinity for technology interaction
7. Social response measures
8. Chatbot use, demographic information
9. Debriefing, final consent, end

Original and Translated Pretest Measures

Open-ended feedback

Was hat Ihnen an der Interaktion mit dem Chatbot (nicht) gefallen? Bitte beschreiben Sie, was Ihnen positiv (negativ) aufgefallen ist [...] *What did you like about the interaction with the chatbot? Please describe what was positive (negative) [...]*

Feedback on technical issues

1. Kam es während Ihres Chats mit dem Bot zu Problemen? [nein; ja, aber ich konnte alle Aufgaben trotzdem lösen; ja, ich konnte eine oder mehrere Aufgaben daher nicht lösen] *Did you encounter any problems during your chat with the bot? [no; yes, but I could solve all tasks anyway; yes, and I could not solve one or more tasks because of this]*
2. Um welche Art von Problem(en) handelte es sich? [es gab ein technisches Problem, z.B. plötzlicher Chatabbruch, eingefrorener Bildschirm, Buttons nicht auswählbar,

¹² Most of Appendix B is included in the appendix of the published manuscript (<https://doi.org/10.1080/10447318.2024.2351707>) or is available as supplementary material on osf (<https://osf.io/j8hde/>).

Texteingabe nicht möglich; der Chatbot hat meine Anfragen teilweise nicht verstanden; der Chatbot hat meine Anfragen teilweise missverstanden; sonstiges Problem] *What kind of problem(s)? [there was a technical problem, e.g., sudden chat abort, frozen screen, buttons not selectable, text input not possible, the chatbot partly did not understand my requests, the chatbot partially misunderstood my requests, other problem]*

3. Wie genau äußerten sich die Probleme? [...] *How exactly did the problems manifest themselves? [...] add*

User Experience Questionnaire Short (UEQ), adapted from Schrepp et al. (2017)

Entscheiden Sie möglichst spontan, welcher der folgenden widersprüchlichen Begriffe den Chatbot besser beschreibt [behindernd–unterstützend, kompliziert–einfach, ineffizient–effizient, verwirrend–übersichtlich, langweilig–spanned, uninteressant–interessant, konventionell–originell, herkömmlich–neuartig]. *Decide as spontaneously as possible which of the following contradictory terms describes the chatbot better [obstructive–supportive, complicated–easy, inefficient–efficient, confusing–clear, boring–exciting, not interesting–interesting, conventional–inventive, usual–leading edge].*

System Usability Scale (SUS), translated and adapted from Bangor et al. (2008)

1. Ich denke, dass ich den Chatbot gerne häufig benutzen würde. *I think I would like to use this chatbot frequently.*
2. Ich fand den Chatbot unnötig komplex. *I found this chatbot unnecessarily complex.*
3. Ich fand den Chatbot einfach zu benutzen. *I thought this chatbot was easy to use.*
4. Ich glaube, ich würde die Hilfe einer technisch versierten Person benötigen, um den Chatbot benutzen zu können. *I think I would need the support of a technical person to be able to use this chatbot.*
5. Ich fand, die verschiedenen Funktionen in diesem Chatbot waren gut integriert. *I thought the various functions in this chatbot were well integrated.*
6. Ich denke, der Chatbot enthielt zu viele Unstimmigkeiten. *I think that there was too much inconsistency in this chatbot.*
7. Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem Chatbot sehr schnell lernen. *I would imagine that most people would learn to use this chatbot very quickly.*
8. Ich fand den Chatbot sehr umständlich zu nutzen. *I found this chatbot very awkward to use.*

9. Ich fühlte mich bei der Benutzung des Chatbots sehr sicher. *I felt very confident using this chatbot.*
10. Ich musste eine Menge lernen, bevor ich anfangen konnte den Chatbot zu verwenden. *I needed to learn a lot of things before I could get going with this chatbot.*

Chatbot Usability Questionnaire (CUQ), translated and adapted from Holmes et al. (2019)

1. Die Persönlichkeit des Chatbots war realistisch. *The personality of the chatbot was realistic.*
2. Der Chatbot wirkte zu roboterhaft. *The chatbot seemed too robotic.*
3. Der Chatbot war bei der Ersteinrichtung einladend. *The chatbot was welcoming during initial setup.*
4. Der Chatbot wirkte sehr unfreundlich. *The chatbot seemed very unfriendly.*
5. Der Chatbot hat seinen Umfang und Zweck gut erklärt. *The chatbot explained its scope and purpose well.*
6. Der Chatbot gab keinen Hinweis auf seinen Zweck. *The chatbot gave no indication as to its purpose.*
7. Der Chatbot war einfach zu navigieren. *The chatbot was easy to navigate.*
8. Bei der Benutzung des Chatbots konnte man leicht verwirrt werden. *It would be easy to get confused when using the chatbot.*
9. Der Chatbot hat mich gut verstanden. *The chatbot understood me well.*
10. Der Chatbot hat viele meiner Eingaben nicht erkannt. *The chatbot failed to recognize a lot of my inputs.*
11. Die Antworten des Chatbots waren nützlich, angemessen und informativ. *Chatbot responses were useful, appropriate, and informative.*
12. Die Chatbot-Antworten waren nicht relevant. *Chatbot responses were not relevant.*
13. Der Chatbot kam gut mit Fehlern oder Irrtümern zurecht. *The chatbot coped well with any errors or mistakes.*
14. Der Chatbot schien nicht in der Lage zu sein, mit Fehlern umzugehen. *The chatbot seemed unable to handle any errors.*
15. Der Chatbot war sehr einfach zu bedienen. *The chatbot was very easy to use.*
16. Der Chatbot war sehr komplex. *The chatbot was very complex.*

Deviations from Preregistration

The wording of hypotheses 1 and 2 was slightly changed:

H1. When users interact with a chatbot primarily via free text (as opposed to buttons), (a) perceived warmth and (b) competence of, (c) satisfaction with, (d) intention to use the chatbot, and (e) emotional connection with the company will be higher.

H2. When a chatbot's responses to errors are social (as opposed to neutral), (a) perceived warmth and (b) competence of, (c) satisfaction with, (d) intention to use the chatbot, and (e) emotional connection with the company will be higher.

For better readability, hypotheses 4 and 5 were split into two sub-hypotheses each:

H4.1. Perceived warmth mediates the positive effects of free text interaction on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

H4.2. Perceived competence mediates the positive effects of free text interaction on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

H5.1. Perceived warmth mediates the positive effects of social responses to errors on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

H5.2. Perceived competence mediates the positive effects of social responses to errors on (a) satisfaction with, (b) intention to use the chatbot, and (c) emotional connection with the company.

Procedure

1. Introduction, informed consent
2. Scenario, instructions, attention check 1
3. Chatbot interaction depending on condition
4. **Manipulation check 1**, attention check 2
5. **Likability, perceived intelligence, satisfaction, intention to use, emotional connection with the company**, perceived anthropomorphism
6. UX questionnaires (same as in the pretest)
7. **Manipulation checks 2-5**, feedback on technical issues (same as in pretest)
8. Chatbot use, demographic information
9. Debriefing, final consent, end

Measures

Core Measures

Likability and perceived intelligence, adapted from Bartneck et al. (2009)

Same as Chapter 2 (see Appendix A).

Satisfaction, translated and adapted from Cheng and Jiang (2020)

1. Ich bin mit dem Chatbot zufrieden.
2. Der Chatbot hat einen guten Job gemacht.
3. Der Chatbot hat das getan, was ich erwartet habe.

Intention to use, translated and adapted from Cheng and Jiang (2020)

1. Ich würde den Chatbot weiterhin nutzen.
2. Ich würde den Chatbot für andere Zwecke als meine derzeitige Nutzung verwenden.
3. Ich würde andere Dienste des Chatbots erkunden als die, die ich derzeit benutze.

Emotional connection with the company, translated and adapted from Christodoulides et al. (2006)

1. Ich habe das Gefühl, dass myclothes.deu sich tatsächlich um mich kümmert.
2. Ich habe das Gefühl, dass myclothes.deu mich wirklich versteht.

Perceived ease of use, translated and adapted from Venkatesh (2000)

1. Meine Interaktion mit dem Chatbot ist klar und verständlich. *My interaction with the chatbot is clear and understandable.*
2. Die Interaktion mit dem Chatbot erfordert nicht viel geistige Anstrengung. *The interaction with the chatbot does not require much mental effort.*
3. Ich finde, dass der Chatbot einfach zu bedienen ist. *I find that the chatbot is easy to use.*
4. Ich finde es einfach, den Chatbot dazu zu bringen, das zu tun, was ich will. *I find it easy to get the chatbot to do what I want.*

Chatbot use frequency

Wie oft haben Sie in den letzten drei Monaten einen Chatbot verwendet? [nie, 1 Mal, 2-3 Mal, öfter als 3 Mal] *How often have you used a chatbot in the last three months? [never, 1 time, 2-3 times, more than 3 times]*

Manipulation checks

1. Wie hoch war das restliche Gutscheinguthaben? [1 €, 10 €, 20 €] *How much was the remaining voucher balance? [1 €, 10 €, 20 €]*
2. Wie oft hat der Chatbot Sie nicht verstanden? Wenn Sie sich nicht an die genaue Zahl erinnern können, versuchen Sie zu schätzen. [...] *How many times did the chatbot not understand you? If you can't remember the exact number, try to guess. [...]*
3. Hat sich der Chatbot bei Ihnen entschuldigt, wenn er eine Anfrage nicht verstanden hat? [nein, ja, weiß nicht] *Has the chatbot apologized to you when it didn't understand a request? [no, yes, don't know]*

4. Hat der Chatbot gesagt, dass er mit jedem Gespräch dazulernt? [nein, ja, weiß nicht]
Did the chatbot say it learns with each conversation? [no, yes, don't know]
5. Hätten Sie das Gespräch zu einem bestimmten Zeitpunkt gerne abgebrochen? [nein, ja]
Would you have liked to have stopped the conversation at any point? [no, yes]

Scenario and Instructions

Original Scenario Description

Zu Ihrem Geburtstag haben Sie einen Geschenkgutschein für den Online-Shop www.myclclothes.deu bekommen. Sie kennen den Shop nicht, aber möchten gerne Ihren Gutschein einlösen. Sie rufen die Seite auf und sehen, dass www.myclclothes.deu Bekleidung und Schuhe für Damen und Herren führt.

Nachdem Sie eine Zeit lang im Shop gestöbert haben, bestellen Sie zwei Paar Sneakers und drei T-Shirts zur Auswahl. Sie bezahlen mit dem Geschenk-Gutschein, den Sie zum Geburtstag bekommen haben. Ihre Bestellung ist nun auf dem Weg zu Ihnen.

Dann entdecken Sie auf www.myclclothes.deu einen Chatbot, der Fragen zum Kundenservice beantworten kann. Ein Chatbot ist ein virtueller Roboter, der mit Ihnen über ein Chatfenster anstelle eines Menschen kommuniziert. Da Sie noch einige Fragen haben, entscheiden Sie sich, den Chatbot zu nutzen. II. Sie möchten folgende Informationen erhalten: Sie möchten Ihr verbleibendes Gutscheinguthaben einsehen. Sie möchten wissen, was passiert, wenn Sie bei der Lieferung nicht zuhause sind. Sie möchten wissen, wo sich Ihr Paket gerade befindet. Sie möchten wissen, wie hoch die Rücksendekosten sind. Optional: Da Sie von den Angeboten begeistert sind, möchten Sie sich gerne zum Newsletter anmelden. III.

Um mithilfe des Chatbots zu den obigen Informationen zu gelangen, benötigen Sie die folgenden Angaben: Der Code auf Ihrer Geschenk-Gutscheinkarte lautet: 3478000022 Der dazugehörige PIN-Code lautet: 2442

Die Sendungsnummer des Pakets lautet: JJD004456789001 Ihre E-Mail-Adresse lautet: a.maier@gmail.de

English Translation of Scenario Description

You have received a gift voucher for the online store www.myclclothes.deu for your birthday. You don't know the store, but would like to redeem your voucher. You open the site and see that www.myclclothes.deu sells clothing and shoes for men and women.

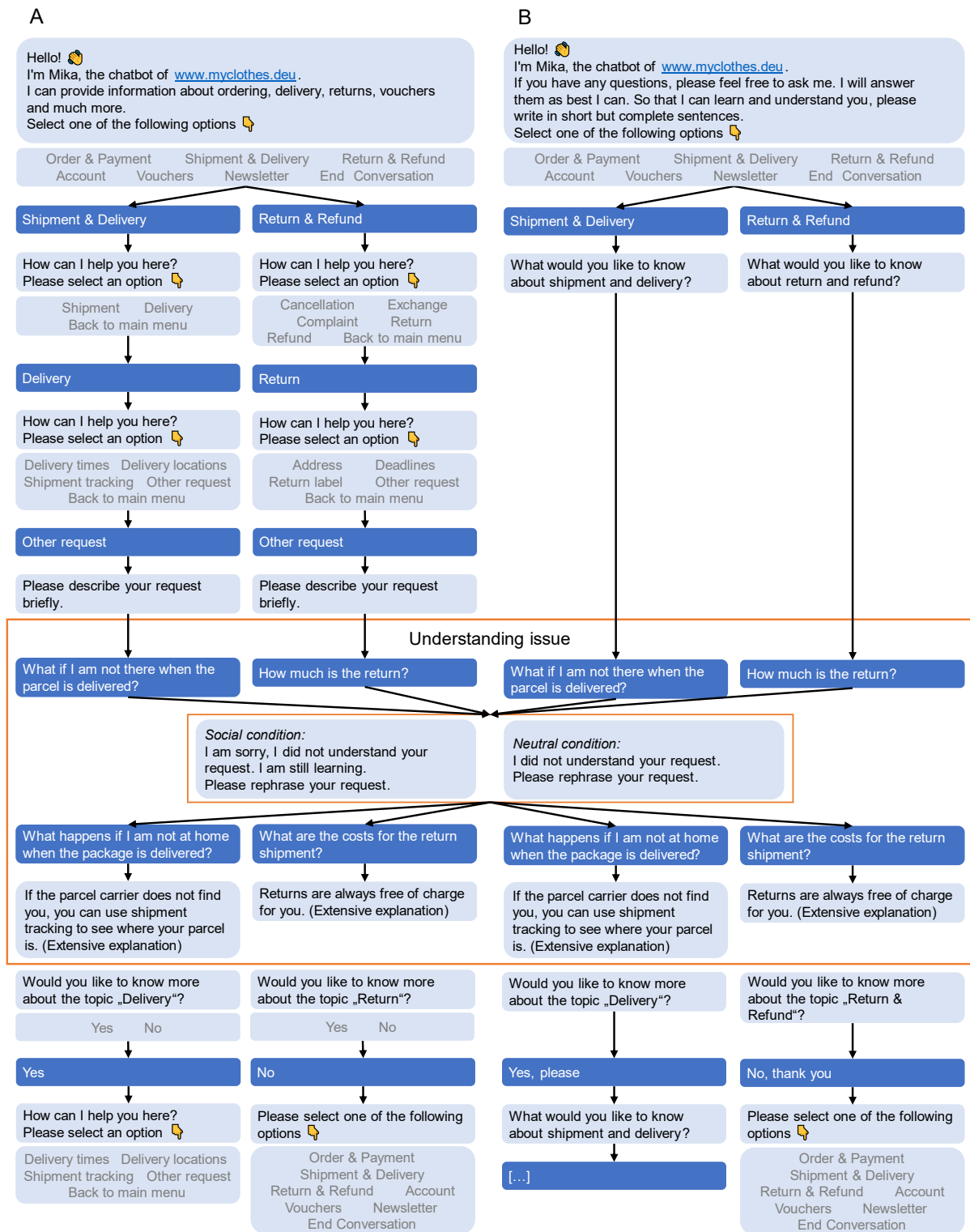
After browsing the store for a while, you order two pairs of sneakers and three t-shirts to choose from. You pay with the gift voucher you received for your birthday. Your order is now on its way to you.

Then you discover a chatbot on www.myclclothes.deu that can answer customer service questions. A chatbot is a virtual robot that communicates with you via a chat window instead of a human being.

Since you still have a few questions, you decide to use the chatbot. you would like to receive the following information: You would like to see your remaining voucher balance. You want to know what happens if you are not at home when the delivery is made. You would like to know where your parcel is currently located. You would like to know how much the return costs are. Optional: As you are enthusiastic about the offers, you would like to subscribe to the newsletter. To access the above information using the chatbot, you will need the following information: The code on your gift voucher card is: 3478000022 The corresponding PIN code is: 2442

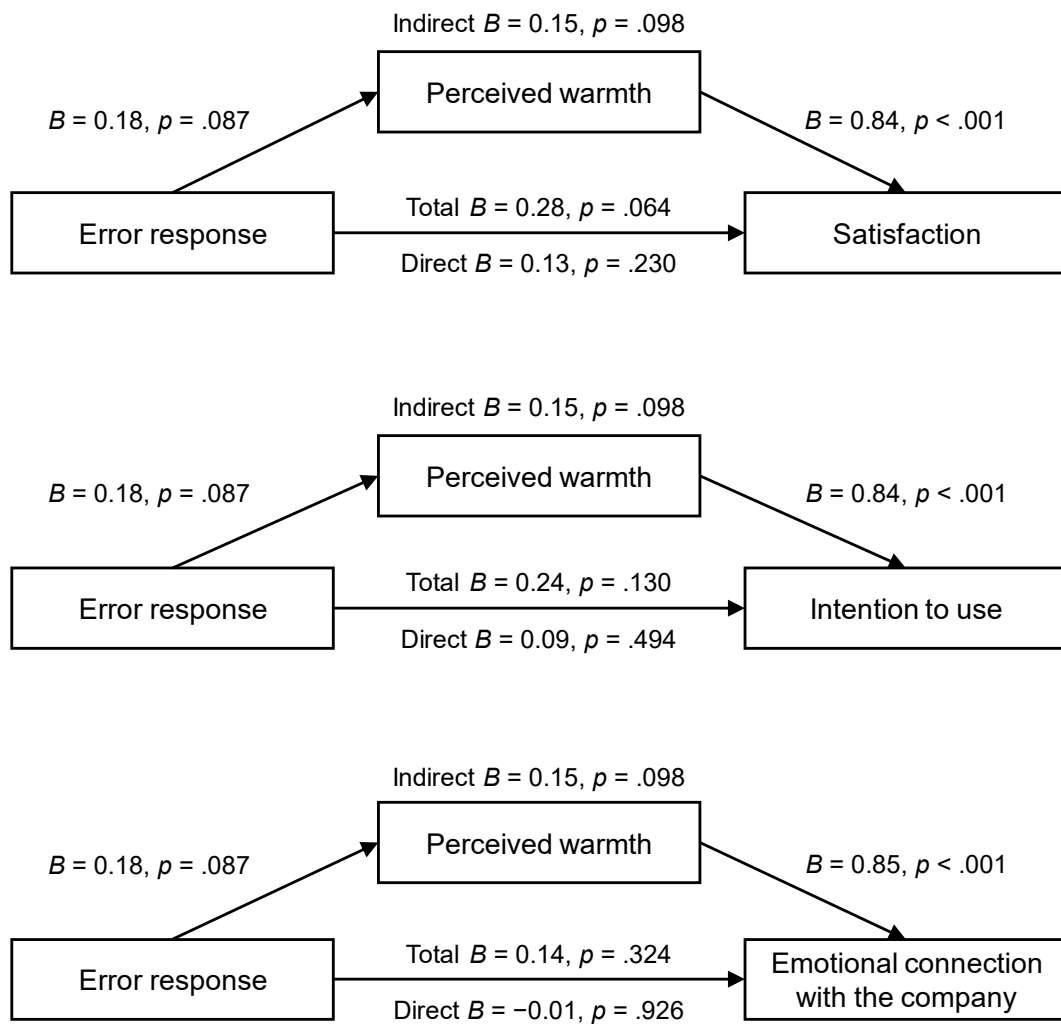
The tracking number of the parcel is: JJD004456789001 Your e-mail address is: a.maier@gmail.de

Figure B.1 Exemplary Translated Conversational Flows in Button and Free Text Chatbot Conditions



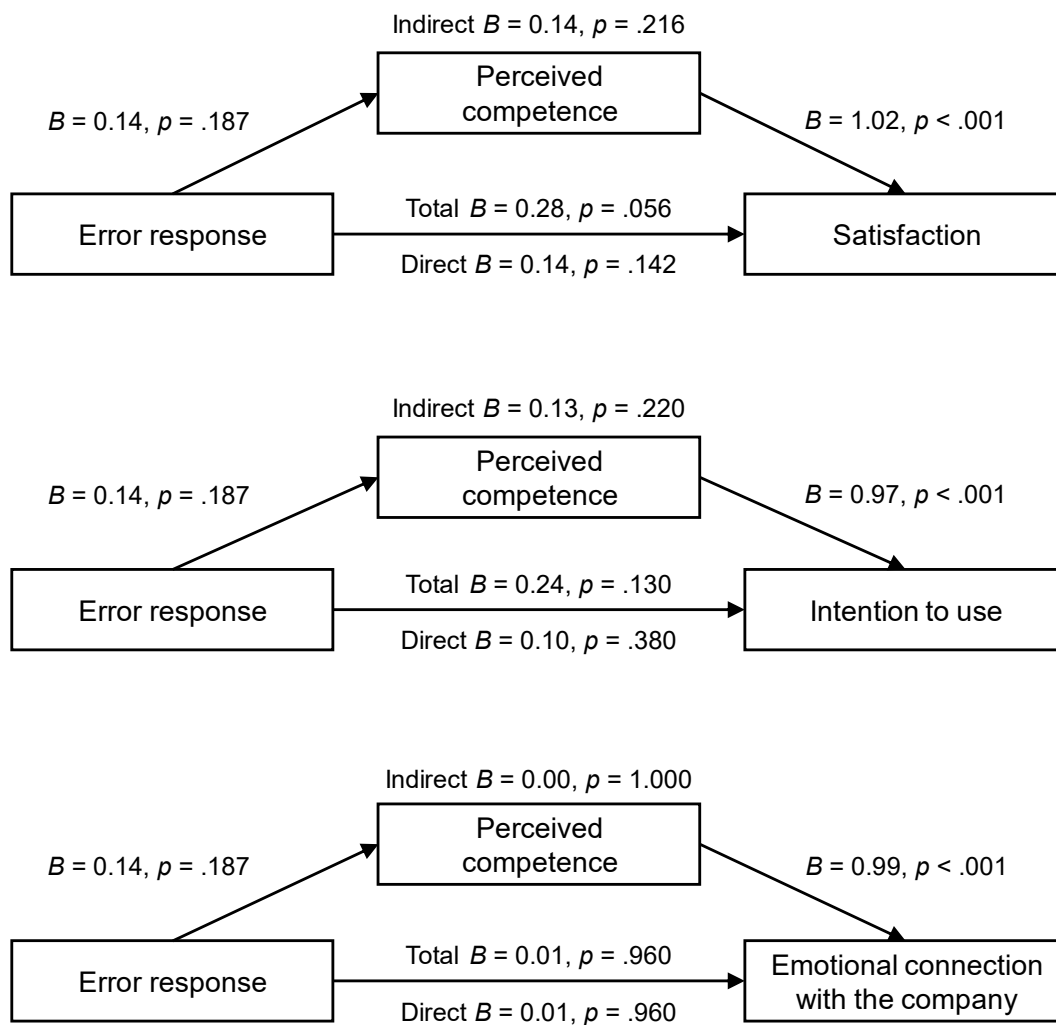
Note. Panels A and B show exemplary conversations with the button and the free text chatbot, respectively. Chatbot responses are light blue, and user responses are dark blue.

Figure B.2 Mediation Results of Error Response on Satisfaction, Intention to Use, and Emotional Connection with the Company via Perceived Warmth



Note. Error response: 0 = neutral and 1 = social. $N = 416$.

Figure B.3 Mediation Results of Error Response on Satisfaction, Intention to Use, and Emotional Connection with the Company via Perceived Competence



Note. Error response: 0 = neutral and 1 = social. $N = 416$.

Table B.1 Results of Exploratory Linear Regression Analyses of Study Variables on Interaction Mechanism and Error Response

Effect	Perceived warmth		Perceived competence		Satisfaction		Intention to use		Emotional connection with the company	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Intercept	5.50***	.10	5.03***	.10	5.30***	.13	4.40***	.15	4.09***	.14
Interaction mechanism	-0.30*	.14	-0.44**	.14	-0.79***	.13	-0.61**	.21	0.01	.20
Error response	0.25°	.15	0.20	.15	0.34°	.20	0.22	.22	0.51*	.21
Interaction mechanism × Error response	-0.12	.20	-0.09	.20	-0.07	.27	0.06	.31	-0.69*	.29

Note. $N = 416$. Interaction mechanism: 0 = buttons and 1 = free text. Error response: 0 = neutral and 1 = social. Unstandardized regression coefficients *B*. Standard errors *SE*.

° $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table B.2 Results of Exploratory Linear Regression Analyses of Study Variables on Interaction Mechanism and Error Response, Controlled for Perceived Ease of Use

Effect	Perceived warmth		Perceived competence		Satisfaction		Intention to use		Emotional connection with the company	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Intercept	3.44***	.19	2.80***	0.18	1.40***	.20	0.82**	.27	1.19***	.27
Perceived Ease of Use	0.40***	.03	0.43***	.18	0.75***	.03	0.69***	.05	0.56***	.05
Interaction mechanism	-0.10	.12	-0.22°	.12	-0.40**	.13	-0.26	.18	0.29°	.17
Error response	0.08	.13	0.02	.12	0.02	.13	-0.07	.18	0.27	.18
Interaction mechanism × Error response	0.01	.17	0.05	.17	0.18	.18	0.28	.25	-0.50*	.25

Note. $N = 416$. Interaction mechanism: 0 = buttons and 1 = free text. Error response: 0 = neutral and 1 = social. Unstandardized regression coefficients *B*. Standard errors *SE*.

° $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table B.3 Results of Exploratory Linear Regression Analyses of Study Variables on Interaction Mechanism and Error Response, Controlled for Previous Chatbot Experience

Effect	Perceived warmth		Perceived competence		Satisfaction		Intention to use		Emotional connection with the company	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Intercept	5.49***	.12	5.06***	.12	5.44***	.16	4.29***	.19	4.01***	.17
Previous experience	0.01	.10	-0.05	.10	-0.22	.14	0.18	.16	0.13	.15
Interaction mechanism	-0.30*	.14	-0.44**	.14	-0.80***	.19	-0.60**	.21	0.01	.20
Error response	0.25 [°]	.15	0.19	.15	0.32 [°]	.20	0.23	.22	0.51*	.21
Interaction mechanism × Error response	-0.12	.20	-0.09	.20	-0.05	.27	0.05	.31	-0.70*	.29

Note. $N = 416$. Interaction mechanism: 0 = buttons and 1 = free text. Error response: 0 = neutral and 1 = social. Unstandardized regression coefficients *B*. Standard errors *SE*.

[°] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

C Appendix Chapter 4

Supplementary information regarding the literature search and screening process, as well as supporting data and the corresponding R code, are accessible at https://osf.io/cezmu/?view_only=48d1e1ccb7b6419294ea54f6fb24cd66.

The preregistration is accessible at https://osf.io/9t2q8/?view_only=b87fe963d36e4ed586865fdfe880a3e3.

Table C.1 Deviations from Protocol

Date	Item	Description	Rationale	Impact
during the writing process	6	The amendment refers to the wording of the research questions. Updated: RQ1. What is the overall effect of text-based chatbots' social cues on users' social responses? RQ2. What factors moderate the effects of text-based chatbots' social cues on users' social responses?	The wording has been adapted to reflect the reference to concepts of the Media Are Social Actors (MASA) paradigm and to text-based chatbots.	No impact.
24/06/2024	7	The amendment refers to a moderator. Updated: Main cues used for manipulation of the independent variable (other vs. verbal) → agent/chatbot category Outdated: Main cues used for manipulation of the independent variable (other vs. verbal) → method category	Whether human-likeness in the agent is manipulated by verbal cues only is a feature of the agent (similar to its avatar).	No impact.
24/06/2024	16	The amendment refers to p curve. Updated: - Outdated: "To assess the evidential value of the findings and to check for selective reporting, we use p-curve (Simonsohn et al., 2014)."	P curve cannot be computed with small sample bias-corrected standardized mean differences (Hedges' g).	No impact because potential publication bias and selective reporting were checked using funnel plots and Egger's regression tests for funnel plot asymmetry.
24/06/2024	15	The amendment refers to the effect size correction. Updated: - Outdated: "The individual effect sizes will be corrected for	A substantial share of studies did not report reliability statistics. In addition, Hedges' g was favored over point-biserial correlation	If at all, the effect sizes are underestimated because they are not corrected (upwards) for measurement error.

Date	Item	Description	Rationale	Impact
24/06/2024	15	<p>measurement error by calculating attenuation factors for each outcome variable per study (Hunter & Schmidt, 2004; Van Berlo et al., 2023).”</p> <p>The amendment refers to the effect size. Updated: “Effect sizes, i.e., standardized mean differences corrected for small sample bias: Hedges’ g), will be computed [...]” Outdated: “Effect sizes, i.e., point-biserial correlations, will be computed [...]”</p>	<p>coefficients. That is why effect sizes were not corrected for measurement error.</p> <p>The computation and interpretation of standardized mean differences is more intuitive as a lot of included studies report Cohen’s d or means and standard deviations to compute Cohen’s d. The calculation of point-biserial correlations would have required a further conversion step.</p>	No differences in overall results are expected.
16/02/2024	8	<p>The amendment refers to the eligibility criterion on Hypotheses. Updated: - Outdated: “Hypotheses. Studies must present one or more hypotheses to test for a change in a user-related outcome resulting from a change in human-likeness or a related variable.”</p>	<p>Only including studies testing hypotheses on changes in a user-related outcome resulting from a change in human-likeness would unnecessarily exclude relevant exploratory studies.</p>	The amendment is necessary because relevant exploratory studies could be overlooked leading to potentially incomplete and biased results.
30/05/2023	11	<p>The correction refers to the first screening of search results. Updated: "Studies will be included 1) if they rely on empirical research, 2) on text-based chatbots, 3) with human-likeness as independent variable, 4) and at least one user-related outcome as dependent variable." Outdated: “Studies will be included if they fulfill the study and report characteristics (see Item 8). [...]”</p>	<p>In the review protocol registered on May 30, 2023, a copy/paste oversight resulted in both screenings referring to all eligibility criteria (Item 8). However, the plan was to consider fewer criteria (in bold) to make the screening process more structured and transparent. The second screening then takes into account all eligibility criteria (Item 8).</p>	The screening process might be less structured and less transparent and the search results might be incomplete which could lead to biases if we do not amend the review protocol.

Table C.2 Literature Search Terms

Title, abstract, keywords		Abstract
Chatbot	Humanlikeness	Empirical study
chatbot* chatterbot* "chat bot*" "chat agent*" "chat assistant*" "AI-based agent*" "AI-based assistant*" "conversation* agent*" "conversation* user interface*" "conversation* interface*" "conversation* UI*" "conversation* system*" "conversation* assistant*" "text-based agent*" "text-based user interface*" "text- based interface*" "text-based UI*" "dialog* system*" "dialog* agent*" "digital assistant*" "intelligent assistant*" "virtual assistant*" "virtual agent*"	anthropomorph* humanlike* human*	empirical* "user stud*" experiment* survey* questionnaire* participant* subject* RCT* study studies

Example Search Query for Web of Science

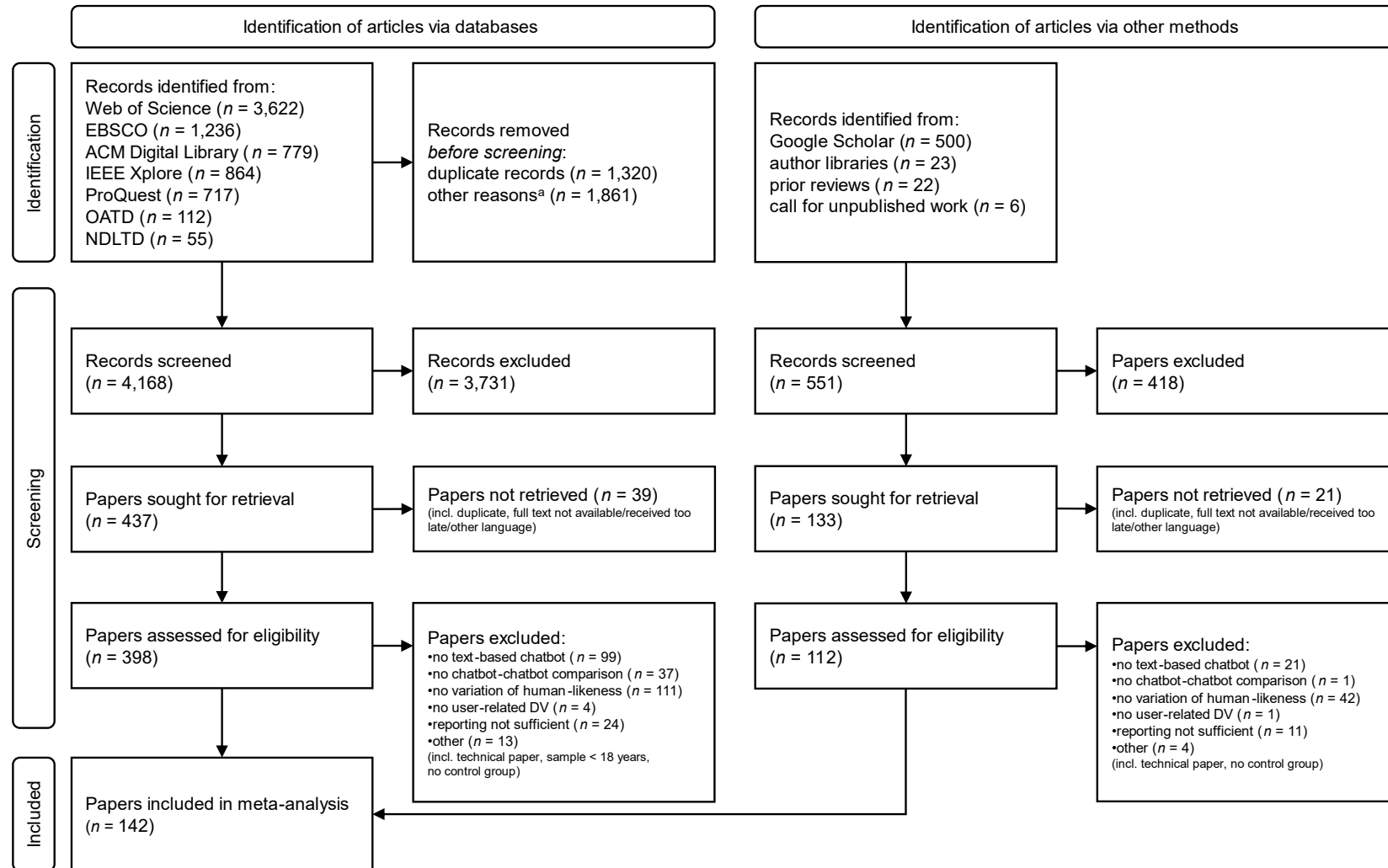
```
((TS=(chatbot* OR chatterbot* OR "chat bot*" OR "chat agent*" OR "chat assistant*"
OR "AI-based agent*" OR "AI-based assistant*" OR "conversation* agent*" OR
"conversation* user interface*" OR "conversation* interface*" OR "conversation* UI*" OR
"conversation* system*" OR "conversation* assistant*"
OR "text-based agent*" OR "text-based user interface*" OR "text-based interface*" OR "text-
based UI*" OR "dialog* system*" OR "dialog* agent*"
OR "digital assistant*" OR "intelligent assistant*" OR "virtual assistant*" OR "virtual
agent*"))
AND TS=(anthropomorph* OR human-like* OR human*))
AND TS=(empirical* OR "user stud*" OR experiment* OR survey* OR questionnaire* OR
participant* OR subject* OR RCT* OR study OR studies)
```

Note. TS searches title, abstract, and author keywords.

Table C.3 Inclusion and Exclusion Criteria

Category	Inclusion criteria	Exclusion criteria
Topic	Text-based chatbots, dialogue systems = text-based software agents interacting with users via natural language (Shawar & Atwell, 2007)	Embodied robots, social robots, including studies on eye gaze, movement, gestures; voice agents/assistants, spoken dialogue systems, speech systems; virtual/embodied/animated avatars; technology paper, e.g., on neural networks, transformer models, machine learning
Study design	Quantitative empirical research, e.g., survey/user study, field experiment, laboratory experiment, online experiment, vignette experiment, quasi-experiment	Non-scientific papers like newspaper/magazine articles, blog entries; qualitative research, case studies; review articles (systematic literature review, meta-analyses etc.); theoretical papers
Sample	Healthy adults where healthy refers to “the absence of disease or impairment” (Sartorius, 2006, p. 662)	Animals, children (< 18), special populations; specific diseases or conditions (e.g., depression, autism)
Language	English or German full text User-related outcome: All constructs/operationalizations regarding attitude toward the chatbot, perception of the chatbot, positive affect, negative affect, user-chatbot relationship, trust, and user behavior; keep all studies that contain some sort of user-related outcome as dependent variable	Other languages
Dependent variable(s)	Human-like, anthropomorphic, social cues, human-likeness, anthropomorphism	Company-related or brand-related outcomes; evaluations of machine learning or other models
Independent variable(s)	Keep all studies that contain some sort of (visual, verbal, chronemic etc.) cue as independent variable	Other independent variables

Figure C.1 The Flow of Papers Throughout the Screening Process



Note. Adapted from Page et al. (2021). ^asee literature search and screening protocol (https://osf.io/cezmu/?view_only=48d1e1ccb7b6419294ea54f6fb24cd66).

Papers Included in the Meta-Analysis

1. Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2), 427–445.
2. Adler, R. F., Iacobelli, F., & Gutstein, Y. (2016). Are you convinced? A wizard of oz study to test emotional vs. rational persuasion strategies in dialogues. *Computers in Human Behavior*, 57, 75–81.
3. Al Farisi, R., Ferdiana, R., & Adji, T. B. (2022). The effect of anthropomorphic design cues on increasing chatbot empathy. *1st International Conference on Information System & Information Technology (ICISIT)*, 370–375.
4. Anonymous. Unpublished manuscript_a.
5. Anonymous. Unpublished manuscript_b.
6. Anonymous. Unpublished manuscript_c.
7. Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85, 183–189.
8. Assink, L. (2019). *Exploring users' perception of chatbots in a mobile commerce environment. Creating a better user experience by implementing anthropomorphic visual and linguistic chatbot features* [Bachelor's thesis, University of Twente].
9. Bae, S., Lee, Y. K., & Hahn, S. (2023). Friendly-bot: The impact of chatbot appearance and relationship style on user trust. *Proceedings of the Annual Meeting of the Cognitive Science Society (COGSCI)*.
10. Bang, J., Han, S., Lee, K., & Lee, G. G. (2015). Open-domain personalized dialog system using user-interested topics in system responses. *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 771–776.
11. Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, 71(3), 409–427.
12. Bergner, A. S., Hildebrand, C., & Häubl, G. (2023). Machine talk: how verbal embodiment in conversational AI shapes consumer–brand relationships. *Journal of Consumer Research*, 50(4), 742–764.
13. Bickmore, T., Schulman, D., & Yin, L. (2009). Engagement vs. deceit: virtual humans with human autobiographies. In Z. Ruttkay, M. Kipp, A. Nijholt, & H. Vilhjalmsson (Eds.), *Intelligent Virtual Agents. IVA 2009. Lecture Notes in Computer Science* (Vol. 5773, pp. 6–19). Springer.
14. Bickmore, T., Schulman, D., & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6), 648–666.

15. Brendel, A. B., Hildebrandt, F., Dennis, A. R., & Riquel, J. (2023). The paradoxical role of humanness in aggression toward conversational agents. *Journal of Management Information Systems*, 40(3), 883–913.
16. Bürke, J., Brendel, A. B., Lichtenberg, S., Diederich, S., & Morana, S. (2021). Do you feel a connection? How the human-like design of conversational agents influences donation behaviour. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation Through Information Systems. WI 2021. Lecture Notes in Information Systems and Organisation* (Vol. 47, pp. 283–298). Springer.
17. Cai, D., Li, H., & Law, R. (2022). Anthropomorphism and OTA chatbot adoption: a mixed methods study. *Journal of Travel & Tourism Marketing*, 39(2), 228–255.
18. Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, 90, 315–330.
19. Chen, J., Guo, F., Ren, Z., Li, M., & Ham, J. (2023). Effects of anthropomorphic design cues of chatbots on users' perception and visual behaviors. *International Journal of Human–Computer Interaction*, 40(14), 3636–3654.
20. Cox, S. R., & Ooi, W. T. (2022). Does chatbot language formality affect users' self-disclosure? *Proceedings of the 4th Conference on Conversational User Interfaces*.
21. Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the bot: anthropomorphism and anger in customer–chatbot interactions. *Journal of Marketing*, 86(1), 132–148.
22. Cui, Y. (Gina). (2022). Sophia Sophia tell me more, which is the most risk-free plan of all? AI anthropomorphism and risk aversion in financial decision-making. *International Journal of Bank Marketing*, 40(6), 1133–1158.
23. De Cicco, R., Silva, S. C. L. da C. e., & Alparone, F. R. (2021). “It’s on its way”: chatbots applied for online food delivery services, social or task-oriented interaction style? *Journal of Foodservice Business Research*, 24(2), 140–164.
24. De Cicco, R., Silva, S.F., Alparone, F.R. (2020). Millennials' attitude toward chatbots: an experimental study in a social relationship perspective. *International Journal of Retail & Distribution Management*. 48(11), 1213–1233.
25. De Sá Siqueira, M. A., Müller, B. C. N., & Bosse, T. (2023). When do we accept mistakes from chatbots? The impact of human-like communication on user experience in chatbots that make mistakes. *International Journal of Human–Computer Interaction*, 40(11), 2862–2872.
26. Diederich, S., Brendel, A. B., Lichtenberg, S., & and Kolbe, L. (2019). Design for fast request fulfillment or natural interaction? Insights from an experiment with a conversational agent. In *Proceedings of the 27th European Conference on Information Systems (ECIS)*.
27. Donkelaar, L. (2018). *How human should a chatbot be? : The influence of avatar appearance and anthropomorphic characteristics in the conversational tone regarding chatbots in customer service field*. [Master's thesis, University of Twente].

28. Duijst, D. (2017). *Can we improve the user experience of chatbots with personalisation?* [Master's thesis, University of Amsterdam].
29. Dybala, P., Ptaszynski, M., Maciejewski, J., Takahashi, M., Rzepka, R., & Araki, K. (2010). Multiagent system for joke generation: humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments*, 2(1), 31–48.
30. Fadhil, A., Schiavo, G., Wang, Y., & Yilma, B. A. (2018). The effect of emojis when interacting with conversational interface assisted health coaching system. In N. Minsky (Ed.), *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 378–383.
31. Farah, J. C., Sharma, V., Ingram, S., & Gillet, D. (2021). Conveying the perception of humor arising from ambiguous grammatical constructs in human-chatbot interaction. *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI)*, 257–262.
32. Gelbrich, K., Hagel, J., & Orsingher, C. (2021). Emotional support from a digital assistant in technology-mediated services: effects on customer satisfaction and behavioral persistence. *International Journal of Research in Marketing*, 38(1), 176–193.
33. Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. *Proceedings of the 26th European Conference on Information Systems (ECIS)*.
34. Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2022). Opposing effects of response time in human-chatbot interaction: the moderating role of prior experience. *Business & Information Systems Engineering*, 64(6), 773–791.
35. 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.
36. Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, 144, 113515.
37. Han, E., Yin, D., & Zhang, H. (2022). Bots with feelings: should AI agents express positive emotion in customer service? *Information Systems Research*, 34(3), 1296–1311.
38. Haugeland, I. K. F., Følstad, A., Taylor, C., & Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: an experimental study of chatbot interaction design. *International Journal of Human-Computer Studies*, 161, 102788.
39. Holtgraves, T. M., Ross, S. J., Weywadt, C. R., & Han, T. L. (2007). Perceiving artificial social agents. *Computers in Human Behavior*, 23(5), 2163–2174.
40. Hu, Y., & Sun, Y. (2023). Understanding the joint effects of internal and external anthropomorphic cues of intelligent customer service bot on user satisfaction. *Data & Information Management*, 7(3), 100047.

41. Huiyang, S., & Min, W. (2022). Improving interaction experience through lexical convergence: the prosocial effect of lexical alignment in human-human and human-computer interactions. *International Journal of Human-Computer Interaction*, 38(1), 28–41.
42. Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2020). Privacy concerns in chatbot interactions. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science* (Vol. 11970, pp. 34–48). Springer.
43. Jakobsen, A. (2021). *Humanlike customer service chatbots: The effect of humanness on user experience and user behavior* [Master's thesis, University of Oslo].
44. Jiang, K., Qin, M., & Li, S. (2022). Chatbots in retail: How do they affect the continued use and purchase intentions of Chinese consumers? *Journal of Consumer Behaviour*, 21(4), 756–772.
45. 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.
46. Jin, S., & Youn, S. (2021). Why do consumers with social phobia prefer anthropomorphic customer service chatbots? Evolutionary explanations of the moderating roles of social phobia. *Telematics and Informatics*, 62, 101644.
47. Jung, J.-Y., Qiu, S., Bozzon, A., & Gadiraju, U. (2022). Great chain of agents: The role of metaphorical representation of agents in conversational crowdsourcing. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–22.
48. Kamoen, N., McCartan, T., & Liebrecht, C. (2022). Conversational agent voting advice applications: A comparison between a structured, semi-structured, and non-structured chatbot design for communicating with voters about political issues. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design. CONVERSATIONS 2021. Lecture Notes in Computer Science* (Vol. 13171, pp. 160–175). Springer.
49. Karma Choedak, K. (2020). *The effect of chatbots response latency on users' trust* [Master's thesis, University of Oklahoma].
50. Kim, W., & Ryoo, Y. (2022). Hypocrisy induction: Using chatbots to promote COVID-19 social distancing. *Cyberpsychology, Behavior, and Social Networking*, 25(1), 27–36.
51. Kim, Y., & Lee, H. (2023). Humanizing chatbots for political campaigns: How do voters respond to feasibility and desirability appeals from political chatbots? *Interacting with Computers*, 35(1), 40–48.
52. Klein, K., & Martinez, L. F. (2022). The impact of anthropomorphism on customer satisfaction in chatbot commerce: An experimental study in the food sector. *Electronic Commerce Research*, 23, 2789–2825

53. Konstanta, S. (2020). *The effects of anthropomorphism and personalization in the context of conversational advertising* [Master's thesis, Copenhagen Business School].
54. 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.
55. Kull, A. J., Romero, M., & Monahan, L. (2021). How may I help you? Driving brand engagement through the warmth of an initial chatbot message. *Journal of Business Research, 135*, 840–850.
56. Kumar, R. (2018). *Socially capable conversational agents for multi-party interactive situations* [Ph.D., Carnegie Mellon University].
57. Kuosmanen, K. (2020). *Purchase intentions on e-commerce: The role of anthropomorphism of disembodied conversational agents* [Master's Thesis, LUT University].
58. Law, E. L.-C., FØLstad, A., & Van As, N. (2022). Effects of humanlikeness and conversational breakdown on trust in chatbots for customer service. *Proceedings of the Nordic Human-Computer Interaction Conference (NordiCHI)*, 1–13.
59. Lee, M., Contreras Alejandro, J., & IJsselsteijn, W. (2024). Cultivating gratitude with a chatbot. *International Journal of Human-Computer Interaction, 40*(18), 4957-4972.
60. Lee, M., Lucas, G., Mell, J., Johnson, E., & Gratch, J. (2019). What's on your virtual mind?: Mind perception in human-agent negotiations. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA)*, 56, 38–45.
61. Lee, S., Lee, E., Park, Y., & Park, G. (2024). Legitimization of paltry favors effect and chatbot-moderated fundraising. *Current Psychology, 43*, 9245–9257.
62. Lee, S., Lee, N., & Sah, Y. J. (2020). Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human-Computer Interaction, 36*(10), 930–940.
63. Lee, S., Park, G., & Chung, J. (2023). Artificial emotions for charity collection: A serial mediation through perceived anthropomorphism and social presence. *Telematics and Informatics, 82*, 102009.
64. Lew, Z., & Walther, J. B. (2023). Social scripts and expectancy violations: Evaluating communication with human or AI chatbot interactants. *Media Psychology, 26*(1), 1–16.
65. Li, J., Zhou, M. X., Yang, H., & Mark, G. (2017). Confiding in and listening to virtual agents: The effect of personality. *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI)*, 275–286.
66. Li, M., & Wang, R. (2023). Chatbots in e-commerce: The effect of chatbot language style on customers' continuance usage intention and attitude toward brand. *Journal of Retailing and Consumer Services, 71*, 103209.
67. Li, Q., Luximon, Y., & Zhang, J. (2023). The influence of anthropomorphic cues on patients' perceived anthropomorphism, social presence, trust building, and acceptance of health care

- conversational agents: Within-subject web-based experiment. *Journal of Medical Internet Research*, 25.
68. Liang, K.-H., Shi, W., Oh, Y., Wang, H.-C., Zhang, J., & Zhou, Y. (2022). *Dialoging resonance: How users perceive, reciprocate and react to chatbot's self-disclosure in conversational recommendations*. arXiv.
 69. Liao, Y. (2021). *Design and evaluation of a conversational agent for mental health support: Forming human-agent sociotechnical and therapeutic relationships* [Ph.D., University of Maryland].
 70. Lichtenberg, S., Bürhke, J., Brendel, A. B., Trang, S., Diederich, S., & Morana, S. (2021). 'Let us work together'— Insights from an experiment with conversational agents on the relation of anthropomorphic design, dialog support, and performance. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Proceedings of the 16th International Conference on Wirtschaftsinformatik* (Vol. 47, pp. 299–315). Springer.
 71. 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.
 72. Liu, C., Zhou, S., Zhang, Y., Liu, D., Peng, Z., & Ma, X. (2022). Exploring the effects of self-mockery to improve task-oriented chatbot's social intelligence. *Designing Interactive Systems Conference*, 1315–1329.
 73. Liu, W. (2019). *Trust in human or machine? Exploring the effect of anthropomorphic cues on user self-disclosure toward a chatbot* [Master's thesis, University of Illinois at Urbana-Champaign].
 74. Liu, Y., Song, C., Huang, D., Liu, S., Li, Z., & Hu, B. (2023). How can chatbots effectively comfort humans? Exploring the interaction of emoji and identity cues for perceived emotional support. *Proceedings of the 32nd European Conference of the International Telecommunications Society (ITS)*, 277999.
 75. Liu-Thompkins, Y., Okazaki, S., & Li, H. (2022). Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *Journal of the Academy of Marketing Science*, 50(6), 1198–1218.
 76. Lou, C., Kang, H., & Tse, C. H. (2022). Bots vs. humans: How schema congruity, contingency-based interactivity, and sympathy influence consumer perceptions and patronage intentions. *International Journal of Advertising*, 41(4), 655–684.
 77. Mai, V., Neef, C., & Richert, A. (2022). "Clicking vs. writing"—The impact of a chatbot's interaction method on the working alliance in AI-based coaching. *Coaching | Theorie & Praxis*, 8(1), 15–31.
 78. Mai, V., Wolff, A., Richert, A., & Preusser, I. (2021). Accompanying reflection processes by an AI-Based StudiCoachBot: A study on rapport building in human-machine coaching using self

- disclosure. In C. Stephanidis, D. Harris, W.-C. Li, D. D. Schmorow, C. M. Fidopiastis, M. Antona, Q. Gao, J. Zhou, P. Zaphiris, A. Ioannou, R. A. Sottilare, J. Schwarz, & M. Rauterberg (Eds.), *HCI International 2021—Late Breaking Papers: Cognition, Inclusion, Learning, and Culture* (Vol. 13096, pp. 439–457). Springer.
79. Meng, J., & Dai, Y. (Nancy). (2021). Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, 26(4), 207–222.
 80. Meng, J., Rheu, M. (Mj), Zhang, Y., Dai, Y., & Peng, W. (2023). Mediated social support for distress reduction: AI chatbots vs. human. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1–25.
 81. Milana, F., Costanza, E., & Fischer, J. E. (2023). Chatbots as advisers: The effects of response variability and reply suggestion buttons. *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI)*, 1–10.
 82. Morana, S., Gnewuch, U., Jung, D., & Granig, C. (2020, June 15). The effect of anthropomorphism on investment decision-making with robo-advisor chatbots. *Proceedings of the 28th European Conference on Information Systems (ECIS)*.
 83. Narducci, F., Basile, P., De Gemmis, M., Lops, P., & Semeraro, G. (2020). An investigation on the user interaction modes of conversational recommender systems for the music domain. *User Modeling and User-Adapted Interaction*, 30(2), 251–284.
 84. Neef, C., Mai, V., & Richert, A. (2022). “I Am scared of viruses, too”—Studying the impact of self-disclosure in chatbots for health-related applications. In M. Kurosu (Ed.), *Human-Computer Interaction. HCI 2022. User Experience and Behavior* (Vol. 13304, pp. 515–530). Springer.
 85. Ng, M., Coopamootoo, K. P. L., Toreini, E., Aitken, M., Elliot, K., & Van Moorsel, A. (2020). Simulating the effects of social presence on trust, privacy concerns & usage intentions in automated bots for finance. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 190–199.
 86. Nguyen, M., Ferm, L., Quach, S., Pontes, N., & Thaichon, P. (2023). Chatbots in frontline services and customer experience: An anthropomorphism perspective. *Psychology & Marketing*, 40(11), 2201–2225.
 87. Ordemann, S., Skjuve, M., & Bjørkli, A. F. and C. A. (2021). “Have a nice day:)”: Social interactions with customer service chatbots. *19th International Conference e-Society 2021*, 53–61.
 88. Park, G., Chung, J., & Lee, S. (2022). Effect of AI chatbot emotional disclosure on user satisfaction and reuse intention for mental health counseling: A serial mediation model. *Current Psychology*, 42, 28663–28673.
 89. Park, G., Chung, J., & Lee, S. (2024). Human vs. machine-like representation in chatbot mental health counseling: The serial mediation of psychological distance and trust on compliance intention. *Current Psychology*, 43, 4352–4363.

90. Park, G., Lee, S., & Chung, J. (2023). Do anthropomorphic chatbots increase counseling satisfaction and reuse intention? The moderated mediation of social rapport and social anxiety. *Cyberpsychology, Behavior, and Social Networking*, 26(5), 357–365.
91. Park, G., Yim, M., Chung, J., & Lee, S. (2022). Effect of AI chatbot empathy and identity disclosure on willingness to donate: The mediation of humanness and social presence. *Behavior & Information Technology*, 42(12), 1998–2010.
92. Pavone, G., Meyer-Waarden, L., & Munzel, A. (2023). Rage against the machine: Experimental insights into customers' negative emotional responses, attributions of responsibility, and coping strategies in artificial intelligence–based service failures. *Journal of Interactive Marketing*, 58(1), 52–71.
93. Peer, W. L., Jr. (2023). *Chatbot pedagogical agent social presence effectiveness with emojis: An experiment* [Ph.D., Grand Canyon University].
94. Pizzi, G., Scarpi, D., & Pantano, E. (2021). Artificial intelligence and the new forms of interaction: Who has the control when interacting with a chatbot? *Journal of Business Research*, 129, 878–890.
95. Pizzi, G., Vannucci, V., Mazzoli, V., & Donvito, R. (2023). I, chatbot! The impact of anthropomorphism and gaze direction on willingness to disclose personal information and behavioral intentions. *Psychology & Marketing*, 40(7), 1372–1387.
96. Ranjartabar, H., Richards, D., Bilgin, A. A., & Kutay, C. (2021). Do you mind if I ask?: Addressing the cold start problem in personalised relational agent conversation. *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents (IVA)*, 167–174.
97. Roy, R., & Naidoo, V. (2021). Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, 126, 23–34.
98. Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, 32(3), 736–751.
99. Scherer, A., Fischer, P. M., Schmitt, B., & Egli, J. (2020). Wow, that’s great! The effect of phatic cues in chatbot conversations [Paper presentation]. *European Marketing Academy Conference (EMAC)*.
100. Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, 37(3), 875–900.
101. Schuetzler, R., Grimes, G., & Giboney, J., (2018). An investigation of conversational agent relevance, presence, and engagement. *AMCIS 2018 Proceedings*, 12.
102. Seeger, A.-M., & Heinzl, A. (2021, June 14). Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service? *ECIS 2021 Research Papers*.

103. Seeger, A.-M., Pfeiffer, J. & Heinzl, A. (2021). Texting with humanlike conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22(4), 931–967.
104. Selamat, M. A., & Windasari, N. A. (2021). Chatbot for SMEs: Integrating customer and business owner perspectives. *Technology in Society*, 66, 101685.
105. Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, 115, 14–24.
106. Shin, H., Bunosso, I., & Levine, L. R. (2023). The influence of chatbot humour on consumer evaluations of services. *International Journal of Consumer Studies*, 47(2), 545–562.
107. Shinozaki, T., Yamamoto, Y., & Tsuruta, S. (2015). Context-based counselor agent for software development ecosystem. *Computing*, 97(1), 3–28.
108. Sidlauskiene, J., Joye, Y., & Auruskeviciene, V. (2023). AI-based chatbots in conversational commerce and their effects on product and price perceptions. *Electronic Markets*, 33(1), 1–21.
109. Sindhu, P., & Bharti, K. (2023). Influence of chatbots on purchase intention in social commerce. *Behaviour & Information Technology*, 43(2), 331–352.
110. Skowron, M., Rank, S., Swiderska, A., Kuster, D., & Kappas, A. (2014). Applying a text-based affective dialogue system in psychological research: Case studies on the effects of system behaviour, interaction context, and social exclusion. *Cognitive Computation*, 6(4), 872–891.
111. Smestad, T. L., & Volden, F. (2019). Chatbot personalities matters. Improving the User Experience of Chatbot Interfaces. In S. S. Bodrunova (Ed.), *Internet science. INSCI 2018. Lecture Notes in Computer Science* (Vol. 11551, pp. 170–181). Springer.
112. Sobel, B. M., & Sims, V. K. (2022). Anthropomorphic perceptions of simple text-based interfaces. In M. Kurosu (Ed.), *Human-Computer Interaction (HCI). User Experience and Behavior* (Vol. 13304, pp. 233–242). Springer.
113. Song, M., Du, J., Xing, X., & Mou, J. (2022). Should the chatbot “save itself” or “be helped by others”? The influence of service recovery types on consumer perceptions of recovery satisfaction. *Electronic Commerce Research and Applications*, 55, 101199.
114. Song, S. W., & Shin, M. (2024). Uncanny effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: The moderating role of avatar familiarity. *International Journal of Human-Computer Interaction*, 40(2), 441–456.
115. Spillner, L., & Wenig, N. (2021). Talk to me on my level – linguistic alignment for chatbots. *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, 45.
116. Srivastava, S., Theune, M., & Catala, A. (2023). The role of lexical alignment in human understanding of explanations by conversational agents. *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI)*, 423–435.
117. Stanciu, G. (2023). *Exploring the interaction between humans and an AI-driven chatbot* [Master’s thesis, University of Oulu].

118. Terada, K., Okazoe, M., Gratch, J. (2021). Effect of politeness strategies in dialogue on negotiation outcomes. *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents (IVA)*, 195–202.
119. Tsumura, T., & Yamada, S. (2023). Influence of agent’s self-disclosure on human empathy. *PLOS ONE*, 18(5), e0283955.
120. Van Hooijdonk, C. (2021). Chatbots in the tourism industry: The effects of communication style and brand familiarity on social presence and brand attitude. *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 375–381.
121. Van Pinxteren, M., Pluymaekers, M., Lemmink, J., & Krispin, A. (2023). Effects of communication style on relational outcomes in interactions between customers and embodied conversational agents. *Psychology & Marketing*, 40(5), 938–953.
122. Wald, R., Heijselaar, E., & Bosse, T. (2021). Make your own: The potential of chatbot customization for the development of user trust. *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 382–387.
123. Wang, J., & Peng, L. (2023). Striking an emotional chord: Effects of emotional appeals and chatbot anthropomorphism on persuasive science communication. *Science Communication*, 45(4), 485–511.
124. Weber, K., Aicher, A., Minker, W., Ultes, S., & André, E. (2023). *Fostering user engagement in the critical reflection of arguments*. arXiv.
125. Weber, U., Lömker, M., & Moskaliuk, J. (2021). The human touch: The impact of anthropomorphism in chatbots on the perceived success of solution focused coaching. *Management Revue*, 32(4), 385–407.
126. Westerman, D., Cross, A. C., & Lindmark, P. G. (2019). I believe in a thing called bot: Perceptions of the humanness of “chatbots.” *Communication Studies*, 70(3), 295–312.
127. Wieland, B., De Wit, J., & De Rooij, A. (2022). Electronic brainstorming with a chatbot partner: A good idea due to increased productivity and idea diversity. *Frontiers in Artificial Intelligence*, 5, 880673.
128. Willems, J., Schmid, M. J., Vanderelst, D., Vogel, D., & Ebinger, F. (2022). AI-driven public services and the privacy paradox: Do citizens really care about their privacy? *Public Management Review*, 25(11), 2116–2134.
129. Xiao, Z., Zhou, M. X., Chen, W., Yang, H., & Chi, C. (2020). If I hear you correctly: Building and evaluating interview chatbots with active listening skills. In R. Bernhaupt (Ed.), *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM.
130. Xu, Y., Bradford, N., & Garg, R. (2023). Transparency enhances positive perceptions of social artificial intelligence. *Human Behavior & Emerging Technologies*, 2023, 5550418.

131. Xu, Y., Zhang, J., Chi, R., & Deng, G. (2022). Enhancing customer satisfaction with chatbots: The influence of anthropomorphic communication styles and anthropomorphised roles. *Nankai Business Review International*, 14(2), 249–271.
132. You, Y., Tsai, C.-H., Li, Y., Ma, F., Heron, C., & Gui, X. (2023). Beyond self-diagnosis: How a chatbot-based symptom checker should respond. *ACM Transactions on Computer-Human Interaction*, 30(4), 1–44.
133. Youn, K., & Cho, M. (2023). Business types matter: New insights into the effects of anthropomorphic cues in AI chatbots. *Journal of Services Marketing*, 37(8), 1032–1045.
134. Youn, S., & Jin, S. V. (2021). “In A.I. we trust?” The effects of parasocial interaction and technopian versus luddite ideological views on chatbot-based customer relationship management in the emerging “feeling economy.” *Computers in Human Behavior*, 119, 106721.
135. Yu, S., Xiong, J. (Jill), & Shen, H. (2024). The rise of chatbots: The effect of using chatbot agents on consumers’ responses to request rejection. *Journal of Consumer Psychology*, 34(1), 35–48.
136. Yu, Z., Black, A. W., & Rudnicky, A. I. (2017). *Learning conversational systems that interleave task and non-task content (version 1)*. arXiv.
137. Yun, J., & Park, J. (2022). The effects of chatbot service recovery with emotion words on customer satisfaction, repurchase intention, and positive word-of-mouth. *Frontiers in Psychology*, 13, 922503.
138. Zhang, T., Feng, C., Chen, H., & Xian, J. (2022). Calming the customers by AI: Investigating the role of chatbot acting-cute strategies in soothing negative customer emotions. *Electronic Markets*, 32(4), 2277–2292.
139. Zhao, R., Romero, O. J., & Rudnicky, A. (2018). SOGO: A social intelligent negotiation dialogue system. *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA)*, 239–246.
140. Zheng, D., Song, R., Hu, T., Fu, H., & Zhou, J. (2020). “Love is as complex as math”: Metaphor generation system for social chatbot. In J.-F. Hong, Y. Zhang, & P. Liu (Eds.), *Chinese Lexical Semantics* (Vol. 11831, pp. 337–347). Springer.
141. Zhou, Y., Fei, Z., He, Y., & Yang, Z. (2022). How human-chatbot interaction impairs charitable giving: The role of moral judgement. *Journal of Business Ethics*, 178(3), 849–865.
142. Zogaj, A., Mähner, P. M., Yang, L., & Tscheulin, D. K. (2023). It’s a match! The effects of chatbot anthropomorphization and chatbot gender on consumer behavior. *Journal of Business Research*, 155, 113412.

Full Results of Subgroup Analyses

Table C.4 Results of Subgroup Analyses for Attitude

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	Δg	<i>SE</i>	<i>Test of Moderators</i>
Sample mean age	124	-0.01	.01	-0.02, 0.01	-1.07	.290	0.01	.01	$F(1, 56) = 1.14, p = .290$
Percentage of female	127	0.01	.005	-0.003, 0.02	1.41	.165	0.01	.01	$F(1, 58) = 1.98, p = .165$
Percentage of previous users	31	-0.01	.01	-0.03, 0.004	-1.60	.137	0.01	.01	$F(1, 11) = 2.57, p = .137$
Assigned chatbot gender							0.06	.17	$F(1, 64) = 0.14, p = .714$
Female	64	0.40	.14	0.12, 0.69	2.82	.006			
Non-female	70	0.34	.11	0.12, 0.56	3.07	.003			
Chatbot avatar							0.11	.16	$F(1, 53) = 0.41, p = .525$
Human-like	84	0.29	.11	0.07, 0.51	2.64	.011			
Not human-like	34	0.19	.14	-0.09, 0.46	1.36	.181			
Chatbot interaction							0.54	.17	$F(1, 51) = 10.18, p = .002$
Unstructured	93	0.57	.10	0.36, 0.78	5.45	< .001			
Structured	43	0.03	.13	-0.24, 0.30	0.20	.843			
Chatbot social cue							0.20	.09	$F(1, 132) = 5.10, p = .026$
Verbal	91	0.35	.10	0.19, 0.50	4.43	< .001			
Other	43	0.15	.08	-0.05, 0.34	1.48	.141			
Chatbot technology							0.03	.54	$F(1, 34) = 0.002, p = .961$
Rule-based	70	0.39	.12	0.14, 0.64	3.22	.003			
Other	4	0.36	.53	-0.70, 1.43	0.69	.491			
Context							0.10	.14	$F(1, 67) = 0.50, p = .481$
Hedonic	39	0.42	.12	0.17, 0.66	3.38	.001			
Functional	99	0.32	.10	0.12, 0.52	3.17	.002			
Task							0.21	.15	$F(1, 67) = 2.15, p = .148$
Critical	48	0.21	.13	-0.06, 0.47	1.56	.123			
Non-critical	90	0.42	.10	0.23, 0.61	4.37	< .001			
Sample							0.17	.14	$F(1, 65) = 1.41, p = .239$
Student	36	0.45	.14	0.18, 0.73	3.26	.002			
Non-student	100	0.28	.10	0.09, 0.48	2.93	.005			
Study setting							0.30	.23	$F(1, 48) = 1.74, p = .193$
Online	112	0.40	.10	0.21, 0.60	4.17	< .001			
Other	17	0.10	.21	-0.32, 0.52	0.49	.624			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Stimulus type							0.02	.19	$F(1, 52) = 0.04, p = .835$
Interaction	91	0.37	.11	0.15, 0.58	3.42	.001			
Prerecorded	47	0.33	.16	0.01, 0.65	2.06	.044			
Scientific field							0.39	.27	$F(1, 52) = 2.04, p = .159$
Communication	19	0.01	.26	-0.50, 0.52	0.05	.960			
Other	119	0.40	.09	0.21, 0.59	4.30	< .001			
Year of publication	122	0.02	.04	-0.06, 0.10	0.49	.626	0.02	.04	$F(1, 49) = 0.24, p = .626$
Sample country							0.06	.18	$F(1, 37) = 0.12, p = .733$
USA	36	0.32	.14	0.04, 0.60	2.34	.024			
Other	60	0.26	.11	0.04, 0.49	2.34	.024			
Publication status							0.44	.37	$F(1, 52) = 1.44, p = .235$
Published	122	0.38	.09	0.20, 0.56	4.21	< .001			
Not published	16	-0.06	.35	-0.77, 0.65	-0.16	.873			
Peer review							0.29	.23	$F(1, 48) = 1.56, p = .218$
Peer reviewed	99	0.41	.10	0.21, 0.61	4.15	< .001			
Not peer reviewed	22	0.12	.21	-0.29, 0.54	0.59	.556			
Preregistration							0.15	.21	$F(1, 67) = 0.50, p = .483$
Preregistered	15	0.49	.21	0.06, 0.92	2.30	.025			
Not preregistered	123	0.35	.09	0.16, 0.53	3.80	< .001			

Note. *k* = number of effect sizes. Figures in bold are significant ($p < .05$).

Table C.5 Results of Subgroup Analyses for Perception

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Sample mean age	289	-0.01	.01	-0.03, 0.004	-1.46	.147	0.01	.01	$F(1, 120) = 2.13, p = .147$
Percentage of female	291	0.01	.01	-0.004, 0.02	1.26	.210	0.01	.01	$F(1, 125) = 1.59, p = .210$
Percentage of previous users	62	0.01	.01	-0.01, 0.03	1.25	.227	0.01	.01	$F(1, 18) = 1.56, p = .227$
Assigned chatbot gender							0.06	.17	$F(1, 122) = 0.13, p = .718$
Female	104	0.49	.14	0.20, 0.77	3.42	< .001			
Non-female	184	0.55	.09	0.36, 0.73	5.80	< .001			
Chatbot avatar							0.07	.19	$F(1, 98) = 0.14, p = .712$
Human-like	157	0.54	.12	0.30, 0.76	4.61	< .001			
Not human-like	87	0.46	.15	0.17, 0.76	3.12	< .001			
Chatbot interaction							0.38	.15	$F(1, 133) = 6.19, p = .014$
Unstructured	205	0.63	.08	0.47, 0.80	7.55	< .001			
Structured	108	0.25	.13	-0.01, 0.51	1.90	.059			
Chatbot social cue							0.21	.13	$F(1, 303) = 2.61, p = .107$
Verbal	235	0.54	.08	0.38, 0.70	6.67	< .001			
Other	70	0.33	.13	0.08, 0.58	2.62	.009			
Chatbot technology							0.30	.32	$F(1, 62) = 0.88, p = .353$
Rule-based	153	0.48	.08	0.31, 0.65	5.61	< .001			
Other	13	0.18	.30	-0.43, 0.79	0.60	.552			
Context							0.01	.14	$F(1, 136) = 0.003, p = .958$
Hedonic	132	0.50	.11	0.29, 0.72	4.55	< .001			
Functional	188	0.51	.09	0.33, 0.69	5.58	< .001			
Task							0.08	.16	$F(1, 136) = 0.27, p = .602$
Critical	76	0.44	.14	0.16, 0.73	3.09	.002			
Non-critical	244	0.53	.08	0.37, 0.69	6.45	< .001			
Sample							0.16	.14	$F(1, 130) = 1.37, p = .245$
Student	54	0.63	.13	0.36, 0.89	4.66	< .001			
Non-student	251	0.46	.08	0.30, 0.63	5.63	< .001			
Study setting							0.19	.20	$F(1, 129) = 0.90, p = .344$
Online	266	0.48	.09	0.31, 0.65	5.55	< .001			
Other	40	0.67	.19	0.30, 1.04	3.57	< .001			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Stimulus type							0.19	.14	$F(1, 136) = 1.74, p = .189$
Interaction	173	0.44	.09	0.26, 0.61	4.87	< .001			
Prerecorded	147	0.63	.11	0.40, 0.85	5.48	< .001			
Scientific field							0.40	.21	$F(1, 97) = 3.76, p = .055$
Communication	61	0.16	.19	-0.22, 0.54	0.82	.412			
Other	259	0.56	.08	0.41, 0.71	7.38	< .001			
Year of publication	281	0.04	.03	-0.02, 0.09	1.38	.170	0.04	.03	$F(1, 94) = 1.92, p = .170$
Sample country							0.20	.18	$F(1, 99) = 1.20, p = .276$
USA	128	0.42	.13	0.17, 0.69	3.25	.001			
Other	124	0.63	.13	0.37, 0.89	4.76	< .001			
Publication status							0.54	.32	$F(1, 97) = 2.86, p = .094$
Published	281	0.54	.07	0.39, 0.68	7.35	< .001			
Not published	39	-0.01	.31	-0.63, 0.61	-0.02	.982			
Peer review							0.23	.25	$F(1, 93) = 0.83, p = .366$
Peer reviewed	255	0.54	.08	0.39, 0.70	6.99	< .001			
Not peer reviewed	25	0.34	.25	-0.16, 0.84	1.34	.183			
Preregistration							0.31	.29	$F(1, 136) = 1.21, p = .273$
Preregistered	39	0.21	.28	-0.34, 0.76	0.77	.443			
Not preregistered	281	0.53	.07	0.38, 0.67	7.18	< .001			

Note. *k* = number of effect sizes. Figures in bold are significant ($p < .05$).

Table C.6 Results of Subgroup Analyses for Positive Affect

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	Δg	<i>SE</i>	<i>Test of Moderators</i>
Sample mean age	21	-0.02	.01	-0.03, 0.001	-1.99	.065	0.02	.01	$F(1, 15) = 3.96, p = .065$
Percentage of female	23	0.01	.01	-0.0001, 0.02	2.09	.052	0.01	.01	$F(1, 17) = 4.37, p = .052$
Percentage of previous users	6	-0.01	.01	-0.04, 0.01	-1.81	.167	0.01	.01	$F(1, 3) = 3.29, p = .167$
Assigned chatbot gender							0.29	.43	$F(1, 15) = 0.45, p = .512$
Female	7	0.14	.39	-0.70, 0.98	0.36	.726			
Non-female	17	0.43	.18	0.04, 0.82	2.36	.032			
Chatbot avatar							0.41	.38	$F(1, 12) = 1.15, p = .305$
Human-like	12	0.53	.27	-0.06, 1.12	1.94	.076			
Not human-like	8	0.12	.27	-0.47, 0.71	0.43	.676			
Chatbot interaction							0.61	.29	$F(1, 17) = 4.55, p = .048$
Unstructured	15	0.64	.17	0.27, 1.00	3.64	.002			
Structured	11	0.02	.23	-0.45, 0.50	0.11	.914			
Chatbot social cue							0.38	.13	$F(1, 23) = 8.67, p = .007$
Verbal	20	0.36	.08	0.19, 0.54	4.33	< .001			
Other	5	-0.02	.13	-0.28, 0.25	-0.14	.887			
Chatbot technology							0.65	.78	$F(1, 15) = 0.71, p = .413$
Rule-based	19	0.43	.17	0.06, 0.79	2.50	.025			
Other	1	1.08	.76	-0.53, 2.69	1.43	.174			
Context							0.55	.29	$F(1, 17) = 3.56, p = .076$
Hedonic	15	0.61	.18	0.24, 0.98	3.47	.003			
Functional	11	0.06	.23	-0.42, 0.55	0.27	.790			
Task							0.24	.25	$F(1, 20) = 0.94, p = .345$
Critical	3	0.20	.26	-0.36, 0.75	0.74	.467			
Non-critical	23	0.44	.15	0.13, 0.76	2.94	.008			
Sample							0.31	.15	$F(1, 20) = 4.42, p = .048$
Student	6	0.66	.19	0.26, 1.05	3.43	.002			
Non-student	20	0.34	.15	0.02, 0.66	2.22	.040			
Study setting							0.14	.36	$F(1, 16) = 0.16, p = .698$
Online	19	0.42	.19	0.03, 0.81	2.26	.038			
Other	6	0.28	.30	-0.37, 0.92	0.91	.375			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Stimulus type							0.15	.26	$F(1, 20) = 0.31, p = .586$
Interaction	21	0.42	.15	0.11, 0.74	2.79	.011			
Prerecorded	5	0.28	.28	-0.31, 0.87	0.98	.341			
Scientific field							0.36	.66	$F(1, 17) = 0.30, p = .592$
Communication	4	0.07	.64	-1.29, 1.42	0.11	.916			
Other	22	0.43	.16	0.10, 0.76	2.72	.015			
Year of publication	22	-0.01	.04	-0.09, 0.06	-0.42	.680	0.01	.04	$F(1, 16) = 0.18, p = .680$
Sample country							0.12	.27	$F(1, 12) = 0.20, p = .660$
USA	5	0.15	.22	-0.34, 0.64	0.68	.508			
Other	15	0.27	.14	-0.04, 0.59	1.90	.082			
Publication status							0.36	.66	$F(1, 17) = 0.30, p = .592$
Published	22	0.43	.16	0.09, 0.76	2.72	.015			
Not published	4	0.07	.64	-1.29, 1.42	0.11	.916			
Peer review							0.07	.70	$F(1, 16) = 0.01, p = .921$
Peer reviewed	21	0.43	.17	0.08, 0.79	2.59	.020			
Not peer reviewed	1	0.36	.68	-1.08, 1.81	0.53	.600			
Preregistration							0.21	.49	$F(1, 17) = 0.18, p = .673$
Preregistered	5	0.22	.46	-0.75, 1.19	0.49	.633			
Not preregistered	21	0.43	.16	0.09, 0.78	2.65	.017			

Note. *k* = number of effect sizes. Figures in bold are significant ($p < .05$).

Table C.7 Results of Subgroup Analyses for Negative Affect

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Sample mean age	31	-0.01	.01	-0.04, 0.01	-1.10	.288	0.01	.01	$F(1, 15) = 1.21, p = .288$
Percentage of female	31	0.01	.01	-0.01, 0.02	0.87	.400	0.01	.01	$F(1, 15) = 0.75, p = .400$
Percentage of previous users	≤ 1								
Assigned chatbot gender							0.17	.38	$F(1, 12) = 0.21, p = .653$
Female	5	0.20	.35	-0.56, 0.97	0.58	.570			
Non-female	26	0.03	.14	-0.28, 0.34	0.21	.838			
Chatbot avatar							0.20	.64	$F(1, 6) = 0.10, p = .767$
Human-like	15	0.05	.23	-0.51, 0.61	0.23	.825			
Not human-like	3	-0.15	.60	-1.62, 1.32	-0.24	.815			
Chatbot interaction							0.31	.26	$F(1, 13) = 1.39, p = .260$
Unstructured	27	0.13	.14	-0.16, 0.43	0.96	.353			
Structured	6	-0.18	.23	-0.67, 0.31	-0.79	.442			
Chatbot social cue							0.13	.17	$F(1, 30) = 0.62, p = .437$
Verbal	26	0.10	.12	-0.14, 0.35	0.85	.400			
Other	6	-0.03	.18	-0.40, 0.34	-0.16	.872			
Chatbot technology							0.11	.57	$F(1, 9) = 0.03, p = .856$
Rule-based	20	0.10	.18	-0.30, 0.50	0.55	.554			
Other	2	-0.01	.54	-1.22, 1.21	-0.01	.990			
Context							0.11	.25	$F(1, 13) = 0.19, p = .666$
Hedonic	7	0.12	.20	-0.31, 0.54	0.60	.561			
Functional	26	0.01	.16	-0.33, 0.35	0.04	.968			
Task							0.04	.26	$F(1, 13) = 0.03, p = .867$
Critical	17	0.02	.21	-0.44, 0.48	0.10	.925			
Non-critical	16	0.06	.15	-0.26, 0.39	0.43	.676			
Sample							0.08	.25	$F(1, 13) = 0.11, p = .747$
Student	19	0.10	.18	-0.30, 0.49	0.51	.937			
Non-student	14	0.01	.17	-0.34, 0.37	0.08	.615			
Study setting							1.25	.41	$F(1, 13) = 9.36, p = .009$
Online	32	-0.03	.10	-0.23, 0.18	-0.28	.783			
Other	1	1.23	.40	0.37, 2.09	3.08	.009			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Stimulus type							0.12	.15	$F(1, 17) = 0.67, p = .424$
Interaction	27	0.07	.12	-0.19, 0.33	0.58	.570			
Prerecorded	6	-0.05	.17	-0.41, 0.31	-0.29	.778			
Scientific field							0.05	.48	$F(1, 13) = 0.01, p = .928$
Communication	4	0.09	.47	-1.91, 1.10	0.20	.848			
Other	29	0.05	.13	-0.23, 0.32	0.36	.722			
Year of publication	33	-0.08	.05	-0.18, 0.02	-1.68	.117	0.08	.05	$F(1, 13) = 2.82, p = .117$
Sample country							0.27	.28	$F(1, 10) = 0.88, p = .371$
USA	14	-0.03	.21	-0.50, 0.45	-0.13	.902			
Other	12	0.24	.19	-0.18, 0.65	1.28	.229			
Publication status									—
Published	33								
Not published	—								
Peer review									—
Peer reviewed	33								
Not peer reviewed	—								
Preregistration									—
Preregistered	—								
Not preregistered	33								

Note. *k* = number of effect sizes. Figures in bold are significant ($p < .05$).

Table C.8 Results of Subgroup Analyses for Rapport

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Sample mean age	86	-0.01	.01	-0.03, 0.004	-1.48	.145	0.01	.01	$F(1, 58) = 2.18, p = .145$
Percentage of female	91	0.01	.005	-0.001, 0.02	1.87	.066	0.01	.005	$F(1, 62) = 3.51, p = .066$
Percentage of previous users	34	0.01	.004	-0.003, 0.01	1.31	.206	0.01	.004	$F(1, 18) = 1.72, p = .206$
Assigned chatbot gender							0.19	.15	$F(1, 53) = 1.71, p = .197$
Female	38	0.53	.12	0.29, 0.77	4.50	< .001			
Non-female	53	0.34	.09	0.17, 0.51	3.94	< .001			
Chatbot avatar							0.16	.18	$F(1, 39) = 0.76, p = .387$
Human-like	48	0.46	.11	0.23, 0.68	4.11	< .001			
Not human-like	24	0.30	.14	0.01, 0.59	2.07	.045			
Chatbot interaction							0.23	.15	$F(1, 55) = 2.54, p = .116$
Unstructured	66	0.46	.08	0.30, 0.62	5.83	< .001			
Structured	30	0.23	.12	-0.02, 0.47	1.85	.069			
Chatbot social cue							0.32	.08	$F(1, 93) = 13.89, p < .001$
Verbal	77	0.45	.07	0.31, 0.59	6.54	< .001			
Other	18	0.13	.09	-0.05, 0.32	1.42	.160			
Chatbot technology							0.26	.27	$F(1, 40) = 0.90, p = .349$
Rule-based	56	0.25	.06	0.13, 0.36	4.37	< .001			
Other	2	-0.01	.27	-0.55, 0.53	-0.04	.971			
Context							0.04	.13	$F(1, 67) = 0.08, p = .778$
Hedonic	39	0.42	.10	0.21, 0.63	4.02	< .001			
Functional	58	0.38	.08	0.21, 0.55	4.51	< .001			
Task							0.23	.15	$F(1, 67) = 2.57, p = .114$
Critical	21	0.23	.13	-0.02, 0.48	1.80	.076			
Non-critical	76	0.46	.08	0.31, 0.61	6.01	< .001			
Sample							0.20	.12	$F(1, 66) = 2.73, p = .103$
Student	34	0.51	.10	0.30, 0.72	4.90	< .001			
Non-student	62	0.31	.08	0.16, 0.46	4.05	< .001			
Study setting							0.04	.17	$F(1, 55) = 0.05, p = .819$
Online	76	0.38	.08	0.23, 0.53	5.01	< .001			
Other	20	0.42	.15	0.12, 0.72	2.84	.006			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Stimulus type							0.48	.14	<i>F</i>(1, 56) = 12.57, <i>p</i> < .001
Interaction	61	0.26	.07	0.12, 0.41	3.65	< .001			
Prerecorded	36	0.75	.12	0.51, 0.98	6.43	< .001			
Scientific field							0.37	.20	<i>F</i> (1, 56) = 3.24, <i>p</i> = .077
Communication	19	0.07	.19	-0.31, 0.46	0.37	.711			
Other	78	0.44	.07	0.30, 0.58	6.36	< .001			
Year of publication	83	-0.01	.04	-0.08, 0.06	-0.26	.793	0.01	.04	<i>F</i> (1, 54) = 0.07, <i>p</i> = .793
Sample country							0.30	.15	<i>F</i>(1, 47) = 4.08, <i>p</i> = .049
USA	32	0.22	.11	-0.01, 0.45	1.89	.065			
Other	55	0.52	.10	0.33, 0.71	5.44	< .001			
Publication status							0.30	.31	<i>F</i> (1, 56) = 0.93, <i>p</i> = .338
Published	83	0.41	.07	0.27, 0.55	6.02	< .001			
Not published	14	0.11	.30	-0.49, 0.72	0.37	.713			
Peer review							0.06	.23	<i>F</i> (1, 54) = 0.06, <i>p</i> = .809
Peer reviewed	77	0.42	.07	0.27, 0.57	5.74	< .001			
Not peer reviewed	6	0.33	.24	-0.14, 0.81	1.40	.167			
Preregistration							0.26	.27	<i>F</i> (1, 56) = 0.96, <i>p</i> = .331
Preregistered	15	0.15	.26	-0.37, 0.67	0.58	.565			
Not preregistered	82	0.41	.07	0.28, 0.55	6.01	< .001			

Note. *k* = number of effect sizes. Figures in bold are significant (*p* < .05).

Table C.9 Results of Subgroup Analyses for Trust

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	Δg	<i>SE</i>	<i>Test of Moderators</i>
Sample mean age	44	0.0002	.01	-0.02, 0.02	0.02	.983	0.0002	.01	$F(1, 22) = 0.0005, p = .983$
Percentage of female	48	-0.01	.01	-0.02, 0.02	-0.11	.913	0.01	.01	$F(1, 26) = 0.01, p = .913$
Percentage of previous users	5	0.02	.02	-0.06, 0.09	0.98	.430	0.02	.02	$F(1, 2) = 0.96, p = .430$
Assigned chatbot gender							0.28	.21	$F(1, 24) = 1.81, p = .191$
Female	14	0.24	.18	-0.12, 0.60	1.34	.194			
Non-female	32	0.52	.12	0.29, 0.76	4.51	< .001			
Chatbot avatar							0.27	.48	$F(1, 18) = 0.31, p = .586$
Human-like	35	0.41	.12	0.17, 0.66	3.51	.002			
Not human-like	2	0.15	.46	-0.83, 1.12	0.32	.755			
Chatbot interaction							0.43	.20	$F(1, 26) = 4.65, p = .040$
Unstructured	33	0.58	.11	0.35, 0.81	5.25	< .001			
Structured	15	0.16	.16	-0.18, 0.49	0.94	.354			
Chatbot social cue							0.39	.21	$F(1, 41) = 3.38, p = .073$
Verbal	30	0.54	.12	0.31, 0.78	4.65	< .001			
Other	13	0.15	.18	-0.20, 0.51	0.87	.391			
Chatbot technology							0.11	.45	$F(1, 20) = 0.06, p = .806$
Rule-based	33	0.44	.13	0.18, 0.71	3.47	.002			
Other	3	0.56	.43	-0.35, 1.46	1.28	.214			
Context							0.30	.20	$F(1, 26) = 2.27, p = .144$
Hedonic	16	0.24	.17	-0.10, 0.58	1.47	.154			
Functional	32	0.55	.12	0.31, 0.79	4.75	< .001			
Task							0.38	.22	$F(1, 26) = 3.03, p = .093$
Critical	11	0.16	.19	-0.24, 0.55	0.82	.419			
Non-critical	37	0.54	.11	0.32, 0.76	5.05	< .001			
Sample							0.10	.23	$F(1, 26) = 0.18, p = .674$
Student	15	0.49	.19	0.11, 0.88	2.63	.014			
Non-student	33	0.39	.14	0.12, 0.67	2.91	.007			
Study setting							0.01	.25	$F(1, 22) = 0.002, p = .969$
Online	34	0.44	.13	0.17, 0.71	3.42	.002			
Other	10	0.43	.22	-0.02, 0.88	1.98	.060			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Stimulus type							0.01	.25	$F(1, 26) = 0.0005, p = .983$
Interaction	36	0.43	.13	0.16, 0.69	3.33	.003			
Prerecorded	12	0.43	.21	-0.01, 0.87	2.02	.054			
Scientific field							0.37	.31	$F(1, 26) = 1.41, p = .246$
Communication	5	0.12	.29	-0.48, 0.72	0.42	.677			
Other	43	0.49	.10	0.28, 0.69	4.86	< .001			
Year of publication	46	0.01	.04	-0.08, 0.10	0.20	.844	0.01	.04	$F(1, 25) = 0.04, p = .844$
Sample country							0.01	.21	$F(1, 21) = 0.001, p = .976$
USA	17	0.31	.15	-0.01, 0.62	2.01	.058			
Other	17	0.30	.14	0.01, 0.59	2.17	.042			
Publication status							0.39	.47	$F(1, 26) = 0.69, p = .414$
Published	46	0.46	.10	0.25, 0.67	4.55	< .001			
Not published	2	0.07	.46	-0.87, 1.01	0.46	.881			
Peer review							0.29	.35	$F(1, 25) = 0.69, p = .414$
Peer reviewed	37	0.53	.11	0.30, 0.76	4.81	< .001			
Not peer reviewed	9	0.21	.22	-0.24, 0.67	0.97	.343			
Preregistration							0.35	.38	$F(1, 26) = 0.83, p = .371$
Preregistered	3	0.12	.37	-0.64, 0.89	0.33	.746			
Not preregistered	45	0.47	.10	0.27, 0.68	4.75	< .001			

Note. *k* = number of effect sizes. Figures in bold are significant ($p < .05$).

Table C.10 Results of Subgroup Analyses for Behavioral Outcomes

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	Δg	<i>SE</i>	<i>Test of Moderators</i>
Sample mean age	124	0.001	.01	-0.01, 0.01	0.12	.907	0.001	.01	$F(1, 67) = 0.01, p = .907$
Percentage of female	126	0.004	.004	-0.003, 0.01	1.12	.267	0.004	.004	$F(1, 70) = 1.25, p = .267$
Percentage of previous users	22	0.01	.01	-0.005, 0.02	1.26	.226	0.01	.01	$F(1, 15) = 1.60, p = .226$
Assigned chatbot gender							0.22	.12	$F(1, 71) = 3.51, p = .065$
Female	41	0.02	.10	-0.18, 0.22	0.22	.829			
Non-female	88	0.25	.07	0.11, 0.38	3.58	< .001			
Chatbot avatar							0.15	.14	$F(1, 43) = 1.08, p = .305$
Human-like	68	0.10	.07	-0.05, 0.25	1.29	.203			
Not human-like	26	0.25	.12	-0.004, 0.50	1.98	.054			
Chatbot interaction							0.20	.11	$F(1, 76) = 3.56, p = .063$
Unstructured	75	0.25	.07	0.12, 0.39	3.83	< .001			
Structured	60	0.05	.08	-0.11, 0.22	0.64	.521			
Chatbot social cue							0.24	.10	$F(1, 129) = 5.47, p = .021$
Verbal	101	0.23	.06	0.11, 0.34	3.91	< .001			
Other	30	-0.01	.10	-0.21, 0.18	-0.13	.895			
Chatbot technology							0.36	.28	$F(1, 49) = 1.66, p = .204$
Rule-based	93	0.19	.06	0.07, 0.31	3.11	.003			
Other	3	0.55	.28	-0.005, 1.10	1.99	.052			
Context							0.15	.11	$F(1, 78) = 1.77, p = .188$
Hedonic	49	0.27	.09	0.09, 0.44	2.99	.004			
Functional	89	0.12	.06	-0.01, 0.25	1.90	.062			
Task							0.08	.73	$F(1, 78) = 0.53, p = .468$
Critical	43	0.11	.09	-0.07, 0.30	1.20	.234			
Non-critical	95	0.20	.06	0.07, 0.32	3.13	.002			
Sample							0.07	.12	$F(1, 77) = 0.28, p = .596$
Student	35	0.12	.11	-0.09, 0.34	1.12	.268			
Non-student	102	0.18	.06	0.06, 0.31	3.01	.004			
Study setting							0.19	.14	$F(1, 77) = 1.93, p = .169$
Online	109	0.20	.06	0.09, 0.32	3.45	< .001			
Other	24	0.01	.12	-0.23, 0.26	0.10	.922			

Moderator	Estimate for moderator level						Moderator analysis		
	<i>k</i>	<i>g</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	$ \Delta g $	<i>SE</i>	<i>Test of Moderators</i>
Stimulus type							0.09	.11	$F(1, 78) = 0.63, p = .431$
Interaction	97	0.20	.06	0.07, 0.32	3.16	.002			
Prerecorded	41	0.11	.09	-0.08, 0.30	1.15	.255			
Scientific field							0.15	.15	$F(1, 67) = 1.11, p = .297$
Communication	18	0.04	.13	-0.23, 0.31	0.28	.780			
Other	120	0.19	.06	0.08, 0.31	3.43	.001			
Year of publication	129	-0.02	.02	-0.06, 0.02	-0.99	.324	0.02	.02	$F(1, 64) = 0.99, p = .324$
Sample country							0.01	.13	$F(1, 52) = 0.004, p = .950$
USA	53	0.20	.09	0.01, 0.37	2.09	.042			
Other	55	0.19	.09	0.02, 0.38	2.21	.032			
Publication status							0.22	.21	$F(1, 67) = 1.17, p = .284$
Published	129	0.19	.05	0.08, 0.29	3.43	.001			
Not published	9	-0.04	.20	-0.43, 0.36	-0.18	.855			
Peer review							0.10	.15	$F(1, 63) = 0.44, p = .508$
Peer reviewed	113	0.17	.06	0.05, 0.29	2.77	.007			
Not peer reviewed	15	0.29	.16	-0.04, 0.62	1.78	.080			
Preregistration							0.11	.20	$F(1, 78) = 0.28, p = .595$
Preregistered	9	0.07	.19	-0.31, 0.45	0.38	.708			
Not preregistered	129	0.18	.05	0.07, 0.29	3.27	.002			

Note. *k* = number of effect sizes. Figures in bold are significant ($p < .05$).

Summary

We are living in a world where we not only interact with other humans but also with artificial agents like chatbots and voice assistants. When communicating with companies in particular, text-based chatbots conveniently answer our questions and complete tasks for us. Although developed using technologies based on artificial intelligence (AI), chatbots often fail to fulfill users' expectations and satisfy users' needs, for example, because they do not understand requests. This gives rise to two questions explored in this dissertation: Do people prefer the service of a human agent or a chatbot? And if organizations decide to use chatbots, how should these agents behave towards the user? Chapter 2 experimentally finds that people would rather consult a human agent than a chatbot to receive study advice; still, the chatbot interaction was perceived as more enjoyable. In addition, human-like communication through responsive verbal cues can increase users' perceptions of an agent's likability and warmth as well as their overall satisfaction with the service encounter. Chapter 3 corroborates these findings in an experiment with a customer service chatbot by showing that human-like error responses improve user satisfaction. However, the appreciation for a human-like free text interaction with a chatbot is diminished by the lower perceived user-friendliness. Finally, Chapter 4 examines the relationship between human-likeness and users' cognitive, affective, and behavioral responses from a meta-perspective. The chapter confirms the findings from Chapters 2 and 3 by demonstrating that human-likeness has a small positive overall effect, which varies depending on the social response in focus and in different circumstances. Taken together, this dissertation contributes to the research field of human-machine communication by enhancing our understanding of users' social responses to text-based service chatbots. The dissertation provides recommendations for organizations that want to use chatbots effectively and responsibly, points out implications for politics and society and opens up an outlook for future research questions.

Zusammenfassung

Wir leben in einer Welt, in der wir nicht nur mit anderen Menschen, sondern auch mit künstlichen Agenten wie Chatbots und Sprachassistenten interagieren. Insbesondere bei der Kommunikation mit Unternehmen beantworten textbasierte Chatbots bequem unsere Fragen und erledigen Aufgaben für uns. Obwohl Chatbots mit Technologien entwickelt werden, die auf künstlicher Intelligenz (KI) basieren, können sie die Erwartungen und Bedürfnisse der Nutzenden oft nicht erfüllen, weil sie beispielsweise Anfragen nicht verstehen. Daraus ergeben sich zwei Fragen, die in dieser Dissertation untersucht werden: Bevorzugen Menschen den Service eines menschlichen Agenten oder eines Chatbots? Und wenn sich Organisationen für den Einsatz von Chatbots entscheiden, wie sollten sie sich gegenüber den Nutzenden verhalten? In Kapitel 2 wurde experimentell festgestellt, dass Menschen lieber einen menschlichen Agenten als einen Chatbot konsultieren würden, um eine Studienberatung zu erhalten; gleichwohl wurde die Chatbot-Interaktion als unterhaltsamer empfunden. Darüber hinaus konnte eine menschenähnliche Kommunikation durch responsive verbale Hinweise die Wahrnehmung der Sympathie und Wärme eines Agenten sowie die Gesamtzufriedenheit der Nutzenden mit dem Service erhöhen. Kapitel 3 bestätigt diese Ergebnisse in einem Experiment mit einem Kundenservice-Chatbot, indem es zeigt, dass menschenähnliche Fehlerreaktionen die Zufriedenheit der Nutzenden verbessern. Die Wertschätzung für eine menschenähnliche Freitextinteraktion mit einem Chatbot wird jedoch durch die geringere wahrgenommene Benutzerfreundlichkeit gemindert. In Kapitel 4 schließlich wird die Beziehung zwischen Menschenähnlichkeit und den kognitiven, affektiven und verhaltensbezogenen Reaktionen der Nutzenden aus einer Meta-Perspektive betrachtet. Das Kapitel bestätigt die Ergebnisse aus den Kapiteln 2 und 3, indem es zeigt, dass Menschenähnlichkeit einen kleinen positiven Gesamteffekt hat, der je nach der im Fokus stehenden sozialen Reaktion und unter verschiedenen Umständen variiert. Insgesamt trägt diese Dissertation zum Forschungsfeld der Mensch-Maschine-Kommunikation bei, indem sie unser Verständnis der sozialen Reaktionen von Nutzenden auf textbasierte Service-Chatbots verbessert. Die Dissertation gibt Empfehlungen für Organisationen, die Chatbots effektiv und verantwortungsbewusst einsetzen wollen, zeigt Implikationen für Politik und Gesellschaft auf und eröffnet einen Ausblick auf zukünftige Forschungsfragen.